

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

LEANDRO ORLANDIN

**Um Estudo sobre a Acessibilidade de Aplicações Móveis: Uma Análise das
Avaliações de Usuários na Google Play Store**

São Paulo

2020

LEANDRO ORLANDIN

**Um Estudo sobre a Acessibilidade de Aplicações Móveis: Uma Análise das
Avaliações de Usuários na Google Play Store**

Projeto de pesquisa para exame de qualificação apresentado à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo como parte dos requisitos para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Orientador: Prof. Dr. Orientador: Marcelo Medeiros Eler

São Paulo

2020

Prof. Dr. -----
Instituição: -----
Presidente

Prof. Dr. -----
Instituição: -----

Prof. Dr. -----
Instituição: -----

Prof. Dr. -----
Instituição: -----

Lista de figuras

| | |
|--|----|
| Figura 1 – Distribuição das avaliações de acessibilidade entre as aplicações da amostra | 46 |
| Figura 2 – Distribuição da proporção entre avaliações de acessibilidade e o total de avaliações de cada aplicação | 47 |
| Figura 3 – Distribuição de tópicos de acessibilidade e palavras-chave nas avaliações de cada aplicação | 49 |
| Figura 4 – Comparação entre as notas associadas a avaliações que contém requisições com as avaliações que contém elogios à acessibilidade das aplicações | 51 |
| Figura 5 – Notas associadas às avaliações que abordam um tópico de acessibilidade (cf. Tabela 8) | 51 |
| Figura 6 – Notas associadas aos comentários que apresentam uma determinada palavra-chave | 52 |

Lista de tabelas

| | |
|--|----|
| Tabela 1 – Diretrizes e requisitos do princípio Perceptível | 14 |
| Tabela 2 – Diretrizes e requisitos do princípio Operável | 15 |
| Tabela 3 – Diretrizes e requisitos do princípio Compreensível | 16 |
| Tabela 4 – Cronograma | 41 |
| Tabela 5 – Estatísticas da amostra das avaliações do <i>Google Play Store</i> | 43 |
| Tabela 6 – Exemplos de palavras-chave utilizadas para selecionar avaliações relacionadas à acessibilidade das aplicações | 44 |
| Tabela 7 – Número de avaliações e aplicações por palavra-chave | 48 |
| Tabela 8 – Distribuição de avaliações de acessibilidade e aplicações por tópico de acessibilidade | 49 |

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 7 |
| 1.1 | <i>Contextualização</i> | 7 |
| 1.2 | <i>Lacuna de Pesquisa</i> | 8 |
| 1.3 | <i>Objetivo</i> | 9 |
| 1.4 | <i>Justificativa e relevância</i> | 10 |
| 1.5 | <i>Metodologia</i> | 10 |
| 1.6 | <i>Organização</i> | 11 |
| 2 | Revisão de Literatura | 12 |
| 2.1 | <i>Conceitos fundamentais</i> | 12 |
| 2.2 | <i>Padrões de Acessibilidade</i> | 13 |
| 2.2.1 | WCAG | 13 |
| 2.2.2 | BBC Accessibility Guidelines | 14 |
| 2.2.3 | eMAG - Modelo de Acessibilidade em Governo Eletrônico | 21 |
| 2.3 | <i>Trabalhos Relacionados</i> | 23 |
| 2.3.1 | <i>User Feedback in the AppStore: An Empirical Study</i> | 24 |
| 2.3.2 | <i>How can i improve my app? Classifying user reviews for software maintenance and evolution</i> | 25 |
| 2.3.3 | <i>Analyzing reviews and code of mobile apps for better release planning</i> | 26 |
| 2.3.4 | <i>SURF: Summarizer of User Reviews Feedback</i> | 28 |
| 2.3.5 | <i>BECLoMA: Augmenting stack traces with user review information .</i> | 29 |
| 2.3.6 | <i>Why People Hate Your App — Making Sense of User Feedback in a Mobile App Store</i> | 29 |
| 2.3.7 | <i>Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews</i> | 30 |
| 2.3.8 | <i>Retrieving and Analyzing Mobile Apps Feature</i> | 31 |
| 2.3.9 | <i>Online Reviews as First Class Artifacts in Mobile App Development</i> | 31 |
| 2.3.10 | <i>User Reviews Matter! Tracking Crowdsourced - Reviews to Support Evolution of Successful Apps</i> | 32 |
| 2.3.11 | <i>Crowdsourcing user reviews to support the evolution of mobile apps</i> | 33 |

| | | |
|----------|--|-----------|
| 2.3.12 | <i>Mobile App Evolution Analysis based on User Reviews</i> | 33 |
| 2.3.13 | <i>Do Android App Users Care about Accessibility? An Analysis of User Reviews on the Google Play Store</i> | 34 |
| 2.3.14 | <i>Accessibility in User Reviews for Mobile Apps: An Automated Detection Approach</i> | 34 |
| 2.4 | <i>Considerações Finais</i> | 34 |
| 3 | Proposta de Pesquisa | 36 |
| 3.1 | <i>Questões de Pesquisa</i> | 36 |
| 3.2 | <i>Atividades Planejadas e Decisões de Projeto</i> | 37 |
| 3.2.1 | Revisão Bibliográfica | 37 |
| 3.2.2 | Seleção de aplicativos móveis | 37 |
| 3.2.3 | Extração das avaliações dos usuários | 38 |
| 3.2.4 | Extração das solicitações de modificações e alterações de código | 39 |
| 3.2.5 | Seleção de avaliações, solicitações de modificações e alterações | 40 |
| 3.2.6 | Análise das avaliações, solicitações de modificações e alterações | 40 |
| 3.3 | <i>Contribuições Esperadas</i> | 40 |
| 3.4 | <i>Cronograma</i> | 41 |
| 4 | Estudo exploratório e atividades já realizadas | 42 |
| 4.1 | <i>Questões de pesquisa</i> | 42 |
| 4.2 | <i>Seleção de aplicações móveis</i> | 42 |
| 4.3 | <i>Extração e seleção das avaliações</i> | 43 |
| 4.4 | <i>Análise dos resultados</i> | 45 |
| 4.5 | <i>Discussão</i> | 53 |
| 4.6 | <i>Publicações</i> | 54 |
| 4.7 | <i>Novas configurações de estudo e atividades realizadas</i> | 54 |
| 5 | Considerações finais | 56 |
| | Referências¹ | 57 |

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

1 Introdução

1.1 Contextualização

Uma importante parcela da população mundial possui algum tipo de deficiência. De acordo com a Organização Mundial da Saúde - OMS (*World Health Organization - WHO*)¹, as estimativas são de que este público seja superior a 1 bilhão de pessoas. Considerando a população brasileira, as estimativas do censo de 2010² são de que aproximadamente 18,8% apresenta algum tipo de deficiência visual, 5,1% deficiência auditiva, 7% deficiência motora e 1,4% deficiência mental ou intelectual. É importante salientar também que esses percentuais se agravam para a população idosa.

As pessoas com deficiência enfrentam barreiras na execução das mais diversas atividades cotidianas, incluindo o uso de serviços e de produtos eletrônicos. Neste contexto, a acessibilidade digital tem o objetivo de remover as barreiras que possam impedir ou dificultar os usuários a perceber, entender e operar produtos digitais de forma completa, segura e autônoma (W3C, 2018b; W3C, 2018a). Embora a acessibilidade digital tenha impacto na usabilidade e na qualidade global de qualquer software (GAY; MILLER; WEST, 2018; ISO, 2011), ela tem o foco nas necessidades específicas de pessoas com deficiência (ISO, 2018).

Muitos avanços foram realizados nesta área ao longo dos anos para promover a acessibilidade digital: a criação de tecnologia assistiva, como leitores de tela e navegação por voz; a criação de padrões e guias de acessibilidade para orientar a produção de conteúdo e aplicações acessíveis, como o WCAG (*Web Content Accessibility Guideline*) do W3C (*The World Wide Web Consortium*); a criação de ferramentas para o desenvolvimento e avaliação de acessibilidade (SILVA; ELER; FRASER, 2018); e a publicação de diversas leis que tornam obrigatório o desenvolvimento de produtos digitais acessíveis na esfera pública e privada (LAZAR, 2019). A acessibilidade digital teve o foco inicial nas páginas e aplicações *web* dado que diversas organizações adotaram esta plataforma para a divulgação de informação e para a realização de negócios, mas o aumento do uso de *smartphones* e de suas aplicações trouxe também o foco para a acessibilidade digital móvel.

¹ <http://documents.worldbank.org/curated/en/665131468331271288/Main-report>

² https://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf

1.2 Lacuna de Pesquisa

Infelizmente, apesar da existência de recursos para o desenvolvimento de aplicações acessíveis, diversos estudos evidenciaram uma falta geral de acessibilidade em aplicações móveis de diversas categorias, tamanhos, complexidade e popularidade (SERRA *et al.*, 2015; Eler *et al.*, 2018; YAN; RAMACHANDRAN, 2019; Vendome *et al.*, 2019; ALSHAYBAN; AHMED; MALEK, 2020; ACOSTA-VARGAS *et al.*, 2020). O mesmo fenômeno tem sido observado em aplicações *web*, e por isso diversos estudos foram conduzidos para entender as razões que explicam a falta de acessibilidade observada. A maioria desses estudos investigou a consciência, o conhecimento e as questões organizacionais que impactam o desenvolvimento de software acessível e constatou que a maioria dos desenvolvedores tem pouco conhecimento sobre acessibilidade, as demandas por manutenção e novos produtos de software são muito altas, os prazos são curtos, não há treinamento adequado e o tema acessibilidade não é uma prioridade das organizações (LAZAR; DUDLEY-SPONAUGLE; GREENIDGE, 2004; FREIRE; RUSSO; FORTES, 2008; Oliveira; Eler, 2017; INAL; RIZVANOĞLU; YESILADA, 2019; BARZILAI-NAHON; BENBASAT; LOU, 2008; PUTNAM *et al.*, 2012). Um estudo recente realizado com mais de 870 desenvolvedores de aplicações móveis no Brasil mostrou um comportamento semelhante nesta plataforma específica.

Se por um lado o conhecimento dos desenvolvedores e o contexto organizacional não favorecem o desenvolvimento de aplicações acessíveis, um outro fator pode fazer a diferença neste cenário: as requisições dos usuários. No contexto do desenvolvimento e evolução de aplicações móveis, é comum que as organizações utilizem avaliações (notas e comentários) realizadas pelos usuários nas lojas de aplicativos para planejar a correção de defeitos e o lançamento de novas funcionalidades, pois isso geralmente resulta em melhores avaliações, o que pode dar destaque e potencializar a instalação por outros usuários (NAYEBI; ADAMS; RUHE, 2016; Palomba *et al.*, 2015; PALOMBA *et al.*, 2018; LI; ZHANG; STEFANIDIS, 2018; Ciurumelea *et al.*, 2017; ORTEGA, 2015).

Considerando que os comentários e as notas recebidas nas avaliações dos usuários são utilizados para corrigir falhas e evoluir aplicações móveis, duas questões de pesquisa são destacadas neste trabalho:

- Os usuários de aplicações móveis mencionam aspectos relacionados à acessibilidade em suas avaliações publicadas nas lojas de aplicativos?
- Os problemas de acessibilidade relatados nas avaliações dos usuários têm impacto na evolução das aplicações móveis?

Diversos estudos sobre avaliações dos usuários em lojas de aplicativos foram realizados para entender as demandas dos usuários e as respostas dos desenvolvedores e das organizações (Iacob; Harrison, 2013; Pagano; Maalej, 2013; IACOB; HARRISON; FAILY, 2014; MCILROY *et al.*, 2016; Di Sorbo *et al.*, 2017; Ciurumelea *et al.*, 2017; ORTEGA, 2015; LI; ZHANG; STEFANIDIS, 2018; Pelloni *et al.*, 2018; Panichella *et al.*, 2015). Os trabalhos encontrados na literatura classificam as avaliações dos usuários de diferentes formas: relato de falha, reclamações em geral, requisição de nova funcionalidade, solução de propostas, problemas com o custo, incompatibilidades, dificuldades com a rede, preocupações com segurança e privacidade, consumo de memória e bateria, elogios, e questões relacionadas à interface com o usuário. Apesar de alguns estudos classificarem avaliações relacionadas à interface com o usuário, nenhuma delas fez uma diferenciação ou um detalhamento sobre os aspectos de acessibilidade abordados pelos usuários, tampouco enfatizaram a evolução da acessibilidade das aplicações.

1.3 Objetivo

O objetivo geral deste projeto de pesquisa é investigar se as avaliações feitas por usuários abordam aspectos relacionados à acessibilidade do produto, e se essas avaliações têm algum efeito na melhoria da acessibilidade da aplicação. Mais especificamente, deseja-se saber:

- A quantidade e a distribuição de avaliações relacionadas à acessibilidade da aplicação
- A diversidade de tipos de violações de acessibilidade mencionadas nas avaliações
- O impacto das questões de acessibilidade nas notas atribuídas pelos usuários
- A quantidade e a distribuição de modificações relacionadas à acessibilidade da aplicação realizadas no código-fonte

1.4 *Justificativa e relevância*

Como as avaliações dos usuários são uma importante ferramenta para direcionar a evolução e a melhoria de aplicações móveis, é importante entender se e como as avaliações publicadas em lojas de aplicativos estão sendo utilizadas para informar aos desenvolvedores e às organizações questões relacionadas à acessibilidade das aplicações avaliadas. Além disso, também é fundamental entender se as questões de acessibilidade abordadas pelos usuários são levadas em consideração na evolução das aplicações avaliadas.

Desta forma, os resultados da investigação aqui proposta podem dar evidências de que as pessoas com deficiência deveriam manifestar explicitamente as barreiras que enfrentam para usar aplicações, e assim pressionar desenvolvedores e organizações a aperfeiçoarem seus produtos. Adicionalmente, mostrar que os diversos tipos de violações de acessibilidade trazem problemas reais para usuários e podem motivar desenvolvedores a levar em consideração as recomendações dos padrões e guias de acessibilidade enquanto projetam as interfaces de suas aplicações.

1.5 *Metodologia*

Os objetivos desta proposta de pesquisa serão alcançados por meio da execução das seguintes atividades:

- Seleção de um conjunto de aplicações móveis cujas avaliações publicadas na loja de aplicativos serão analisadas.
- Obtenção das avaliações (comentários e notas atribuídas) dos usuários das aplicações selecionadas.
- Obtenção das sugestões de modificações (correção de defeitos, melhorias, novas funcionalidades - *issues*) e alterações (*commits*) no código das aplicações selecionadas por meio do acesso a seus repositórios de código.
- Seleção das avaliações dos usuários, sugestões de modificações e alterações que abordam aspectos da acessibilidade das aplicações selecionadas.
- Análise das avaliações, sugestões de modificações e alterações relacionadas à acessibilidade das aplicações.

A metodologia deste trabalho está detalhada no Capítulo 3, onde todas as decisões tomadas para a condução deste projeto de pesquisa são apresentadas e justificadas.

1.6 Organização

Esta proposta de pesquisa está organizada da seguinte forma: no Capítulo 2 são apresentados os conceitos fundamentais que embasam este projeto de pesquisa, no Capítulo 3 a metodologia deste projeto de pesquisa é apresentada em detalhes, e no Capítulo 4 as atividades já realizadas desta proposta de pesquisa são apresentadas. Por fim, algumas considerações finais são apresentadas no Capítulo 5.

2 Revisão de Literatura

Neste capítulo são apresentados alguns conceitos fundamentais de acessibilidade, bem como a descrição dos padrões de acessibilidade utilizados como referência para a condução dos estudos apresentados nesta proposta de pesquisa. Além disso, é apresentada uma visão geral dos trabalhos relacionados.

2.1 *Conceitos fundamentais*

Acessibilidade digital refere-se à capacidade de um produto ou conteúdo digital de ser percebido, compreendido e operado de forma completa, autônoma e segura por qualquer usuário, independentemente de suas capacidades físicas, mentais ou intelectuais (W3C, 2018a). Ainda, refere-se à flexibilidade do software em se adaptar às necessidades de cada usuário, suas preferências e suas limitações (SILVA, 2020). Embora a acessibilidade digital esteja relacionada mais especificamente ao atendimento das necessidades específicas de usuários com deficiência, ela impacta a usabilidade global de qualquer software.

Diversos avanços foram alcançados nas últimas décadas nesta área. Um dos avanços que se destaca é a criação de tecnologias assistivas. As tecnologias assistivas ampliam as capacidades dos usuários para que eles possam compreender e operar um produto digital. Por exemplo, um leitor de tela que enuncia para o usuário com deficiência visual todo o conteúdo textual presente na interface do software; ou um sistema que executa as ações dos usuários utilizando apenas interação por voz ao invés do uso de periféricos como *mouse* e teclado, ou até mesmo toques na tela.

Entretanto, o uso isolado de tecnologias assistivas não são suficientes para tornar um software acessível, pois é preciso que este tenha sido desenvolvido para ser compatível com as diversas tecnologias assistivas existentes. Por exemplo, as aplicações precisam fornecer um conteúdo textual para os elementos não-textuais (como as imagens) para que o leitor de tela possa informar ao usuário qual é a função daquele elemento da interface, seja ela uma possível ação ou uma informação.

Identificar os requisitos necessários para implementar uma interface acessível para os mais diversos tipos de deficiência e as mais diversas tecnologias assistivas é uma tarefa de alto custo. Portanto, diversos padrões e guias de acessibilidade foram criados para

direcionar a criação de software e conteúdos digitais acessíveis. Esses padrões apresentam recomendações ou requisitos que precisam ser implementados para que o software seja acessível ao maior número de usuários possíveis. Os detalhes de alguns desses padrões são apresentados na próxima seção.

2.2 Padrões de Acessibilidade

Os padrões de acessibilidade digital visam a fornecer requisitos ou recomendações para a implementação de um software acessível. Alguns deles tem o foco em conteúdos específicos para aplicações Web, outros para aplicações móveis, e outros combinam elementos dessas duas plataformas. Embora alguns padrões sejam específicos para um tipo de plataforma, os princípios apresentados podem ser adaptados e aplicados em diferentes contextos. A seguir são apresentados os detalhes de três padrões: o WCAG (W3C, 2018b), o padrão da BBC (BBC, 2017) e o e-MAG (GOVERNMENT, 2014).

2.2.1 WCAG

O *Web Content Accessibility Guidelines*, ou WCAG2.1 ¹, é um documento elaborado pela *Web Accessibility Initiative* (WAI) da *The World Wide Web Consortium* (W3C). Este padrão apresenta suas recomendações em formato de declarações testáveis que não se referem a uma tecnologia específica e possui a intenção de tornar o conteúdo *web* mais acessível, com a consciência de que não é capaz de abordar as necessidades de pessoas com todos os tipos, graus e combinações de deficiência. Inicialmente, o WCAG tinha o foco apenas em conteúdos e em aplicações digitais, mas em sua nova versão foram incorporados elementos relacionados à plataforma móvel.

As recomendações do WCAG são classificadas em três níveis de conformidade: A (o mais baixo), AA e AAA (o mais elevado). Além disso, as recomendações estão divididas em quatro princípios fundamentais: *perceptível*, pois a informação deve ser apresentada na interface de forma a garantir que qualquer usuário possa percebê-la; *operável*, pois todos os componentes de navegação da aplicação devem ser operáveis por qualquer usuário; *compreensível*, pois todo usuário deve conseguir entender a informação passada pela interface; *robusto*, pois a interface deve ser robusta o suficiente para que

¹ <https://www.w3.org/WAI/intro/wcag>

Tabela 1 – Diretrizes e requisitos do princípio Perceptível

| ID | DIRETRIZ | REQUISITO |
|--------|------------------------|---|
| 1.1.1 | Alternativas de texto | Conteúdo Não Textual |
| 1.2.1 | Mídia baseada no tempo | Apenas Áudio e apenas Vídeo (Pré-gravado) |
| 1.2.2 | Mídia baseada no tempo | Legendas (Pré-gravado) |
| 1.2.3 | Mídia baseada no tempo | Audiodescrição ou Mídia Alternativa (Pré-gravado) |
| 1.2.4 | Mídia baseada no tempo | Legendas (Ao Vivo) |
| 1.2.5 | Mídia baseada no tempo | Audiodescrição (Pré-gravado) |
| 1.2.6 | Mídia baseada no tempo | Linguagem de sinais (pré-gravada) |
| 1.2.7 | Mídia baseada no tempo | Audiodescrição estendida (pré-gravado) |
| 1.2.8 | Mídia baseada no tempo | Mídia Alternativa (Pré-gravado) |
| 1.2.9 | Mídia baseada no tempo | Apenas áudio (Ao Vivo) |
| 1.3.1 | Adaptável | Informações e Relações |
| 1.3.2 | Adaptável | Sequência com Significado |
| 1.3.3 | Adaptável | Características Sensoriais |
| 1.3.4 | Adaptável | Orientação |
| 1.3.5 | Adaptável | Identificar o objetivo de entrada |
| 1.3.6 | Adaptável | Identificar o propósito |
| 1.4.1 | Distinguível | Utilização de Cores |
| 1.4.2 | Distinguível | Controle de Áudio |
| 1.4.3 | Distinguível | Contraste (Mínimo) |
| 1.4.4 | Distinguível | Redimensionar texto |
| 1.4.5 | Distinguível | Imagens de Texto |
| 1.4.6 | Distinguível | Contraste (Melhorado) |
| 1.4.7 | Distinguível | Som baixo ou sem som de fundo |
| 1.4.8 | Distinguível | Apresentação Visual |
| 1.4.9 | Distinguível | Imagens de texto (sem exceção) |
| 1.4.10 | Distinguível | Reflow |
| 1.4.11 | Distinguível | Contraste sem texto |
| 1.4.12 | Distinguível | Espaçamento de texto |
| 1.4.13 | Distinguível | Conteúdo sobre o foco do mouse ou do teclado |

possa ser interpretada por qualquer usuário, utilizando ou não tecnologias assistivas. As Tabelas 1, 2 e 3 mostram as recomendações para cada princípio do WCAG.

2.2.2 *BBC Accessibility Guidelines*

A BBC - (*British Broadcasting Corporation*) (BBC, 2017) é uma empresa pública líder mundial em serviços de transmissão, incluindo canais de televisão e rádio, bem como transmissões via internet. Com a intenção de manter seus conteúdos disponíveis para o maior número de pessoas, foram definidas diretrizes que devem ser seguidas para os conteúdos disponibilizados pela empresa, e que se tornaram uma das referências globais. O

Tabela 2 – Diretrizes e requisitos do princípio Operável

| ID | DIRETRIZ | REQUISITO |
|--------|------------------------|-------------------------------------|
| 2.1.1 | Operável | Teclado |
| 2.1.2 | Operável | Sem Bloqueio do Teclado |
| 2.1.3 | Operável | Teclado (sem exceção) |
| 2.1.4 | Operável | Atalhos para teclas de caracteres |
| 2.2.1 | Tempo suficiente | Ajustável por limite de tempo |
| 2.2.2 | Tempo suficiente | Colocar em Pausa, Parar, Ocultar |
| 2.2.3 | Tempo suficiente | Sem limite de tempo |
| 2.2.4 | Tempo suficiente | Interrupções |
| 2.2.5 | Tempo suficiente | Re-autenticação (nova autenticação) |
| 2.2.6 | Tempo suficiente | Tempos Limite |
| 2.3.1 | Convulsões | Três Flashes ou Abaixo do Limite |
| 2.3.2 | Convulsões | Três flashes |
| 2.3.3 | Convulsões | Animacão de Interações |
| 2.4.1 | Navegável | Ignorar Blocos |
| 2.4.2 | Navegável | Página com Título |
| 2.4.3 | Navegável | Ordem do Foco |
| 2.4.4 | Navegável | Finalidade do Link (Em Contexto) |
| 2.4.5 | Navegável | Várias Formas (múltiplos caminhos) |
| 2.4.6 | Navegável | Cabeçalhos e Rótulos |
| 2.4.7 | Navegável | Foco Visível |
| 2.4.8 | Navegável | Localização |
| 2.4.9 | Navegável | Objetivo do link (apenas link) |
| 2.4.10 | Navegável | Cabeçalhos das Seções |
| 2.5.1 | Modalidades de Entrada | Gestos de Ponteiro |
| 2.5.2 | Modalidades de Entrada | Cancelamento de ponteiro |
| 2.5.3 | Modalidades de Entrada | Rótulo em nome |
| 2.5.4 | Modalidades de Entrada | Atuação do movimento |
| 2.5.5 | Modalidades de Entrada | Tamanho do Alvo |
| 2.5.6 | Modalidades de Entrada | Mecanismos de Entrada Simultâneos |

conjunto de práticas recomendadas podem ser aplicadas a conteúdos Web móvel, aplicações híbridas e nativas. As diretrizes do padrão da BBC estão divididas em 11 categorias. A seguir é apresentada uma visão geral de cada uma das diretrizes deste padrão e organizadas de acordo com cada categoria:

- **Áudio e Vídeo:**

- Alternativas para conteúdos de vídeo e áudio: quando possível, devem ser disponibilizados conteúdos alternativos, como por exemplo em legendas, linguagem de sinais e transcrições, incorporados à mídia.

Tabela 3 – Diretrizes e requisitos do princípio Compreensível

| ID-DIR | DIRETRIZ | REQUISITO |
|--------|------------------------|---|
| 3.1.1 | Legível | Idioma da Página |
| 3.1.2 | Legível | Idioma das Partes |
| 3.1.3 | Legível | Palavras incomuns |
| 3.1.4 | Legível | Abreviações |
| 3.1.5 | Legível | Nível de Leitura |
| 3.1.6 | Legível | Pronúncia |
| 3.2.1 | Previsível | Em Foco |
| 3.2.2 | Previsível | Em Entrada |
| 3.2.3 | Previsível | Navegação Consistente |
| 3.2.4 | Previsível | Identificação Consistente |
| 3.2.5 | Previsível | Alteração no pedido |
| 3.3.1 | Assistência de entrada | Identificação do Erro |
| 3.3.2 | Assistência de entrada | Rótulos ou Instruções |
| 3.3.3 | Assistência de entrada | Sugestão de Erro |
| 3.3.4 | Assistência de entrada | Prevenção de Erros (Legal, Financeiro, Dados) |
| 3.3.5 | Assistência de entrada | Ajuda disponível |
| 3.3.6 | Assistência de entrada | Prevenção de Erros (Todos) |
| 4.1.1 | Compatível | Análise (código) |
| 4.1.2 | Compatível | Nome, Função, Valor |
| 4.1.3 | Compatível | Mensagens de Status |

- Autoplay: elementos de áudio não devem iniciar automaticamente, exceto se o usuário tiver conhecimento prévio ou que sejam fornecidos botões de controle (pausa, parada e mudo).

- Metadados: Os metadados relevantes devem ser disponibilizados para todas as mídias.

- Controle de Volume: em havendo música de fundo, sons de ambiente ou efeitos sonoros narrativos e de edição, deverão ser disponibilizados controles de volume separadamente.

- Conflito de áudio: não devem ocorrer conflitos dos áudios da tecnologia de assistência nativa com os sons narrativos em jogos ou mídias interativas.

- *Design:*

- Contraste de cores: deve haver um contraste mínimo entre a cor do texto e a cor do conteúdo do plano de fundo.

- Cor e significado: informação ou significado não deve ser transmitido apenas através da diferença de cores.

- Estilo e leitura: o conteúdo principal deve ser acessível, mesmo quando o estilo não for suportado ou for removido intencionalmente pelo usuário.
 - Tamanho de itens clicáveis: devem ser grandes o suficiente para que o usuário possa tocar com precisão.
 - Espaçamento: deve haver um espaço inativo mínimo entre os itens clicáveis.
 - Conteúdo dimensionável: o usuário deve poder controlar o dimensionamento da fonte e a escala (*zoom*) da tela de interface do usuário.
 - Elementos acionáveis: deve ser possível a distinção clara de *links* e outros elementos acionáveis.
 - Foco: todos os elementos devem continuar visíveis, independentemente do foco utilizado pelo usuário.
 - Consistência: a experiência do usuário deve ser consistente, manter a mesma lógica e linguagem, o que facilitará ao usuário prever a próxima etapa.
 - Escolha: as interfaces do usuário devem prover múltiplas formas de interação com o seu conteúdo.
 - Ajustabilidade: mídias interativas, incluindo jogos, devem ser ajustáveis pelo usuário de acordo com sua capacidade (habilidade) e preferência.
 - Cintilação: não deve haver conteúdo que pisque mais do que 3 vezes em um período de 1 segundo. Este tipo de cintilação pode causar fadiga ocular, tonturas, dores de cabeça, enxaqueca, náusea, podendo chegar a vertigem e até mesmo convulsões em casos específicos.
- Editorial:
 - Títulos consistentes: devem ser utilizados em *web sites* e aplicações nativas, de forma a facilitar o entendimento do conteúdo completo, tornando-o familiar, principalmente para usuários que se utilizam de leitores de tela.
 - Indicação do idioma: deve estar especificado o idioma de uma página ou aplicação, sendo que alterações no conteúdo devem ser indicadas ao usuário.
 - Instruções: quando necessário, instruções adicionais devem ser fornecidas de forma a auxiliar o conteúdo disponibilizado em formato de áudio e vídeo.
 - Foco:
 - Elementos focáveis: todos os elementos interativos devem ser focáveis. Elementos inativos não devem ser focáveis.

- Armadilhas de teclado: não deve existir armadilhas de teclado, de forma que o usuário possa controlar a interface através de um teclado ou mesmo uma entrada que não possua um "ponteiro".
 - Ordem do conteúdo: a ordem do conteúdo deve ser lógica, facilitando o entendimento do mesmo principalmente por usuários que se utilizam de tecnologia assistiva.
 - Ordem do foco: o conteúdo clicável deve ser navegável em uma sequência entendível. Por exemplo: navegar em um formulário sem ordem lógica do foco tornará o mesmo desorientador para um usuário com leitor de tela.
 - Interações do usuário: Ações devem desencadear outra interação apropriada, de acordo com o método de entrada de dados pelo usuário, como por exemplo mouse, teclado ou mesmo outros controladores.
 - Métodos de entrada alternativos: Devem ser suportados métodos de entrada alternativos, como por exemplo telas em braille ou simplesmente um teclado.
- Formulários:
 - Rótulos dos controles dos formulários: todos os controles dos formulários devem possuir rótulos exclusivos e disponíveis para tecnologias assistidas, facilitando o entendimento.
 - Entrada de dados: deve ser claramente indicado e suportado um formato de entrada de dados padrão, facilitando o usuário de entender e acertar a entrada na primeira vez.
 - *Layout* dos formulários: os rótulos devem ser colocados próximos dos controles do formulário, reduzindo o risco do usuário de se desorientar.
 - Agrupamento de elementos: os controles, rótulos e outros elementos do formulário devem estar adequadamente agrupados, o que reduzirá o número de passos e complexidade de preenchimento principalmente por usuários que se utilizam de tecnologia assistiva.
 - Foco manuseável: O foco ou o contexto não devem mudar automaticamente durante a entrada de dados, mas sim apenas com uma ação do próprio usuário.
 - Imagens:
 - Imagens de texto: devem ser evitadas, já que se trata de uma forma inflexível de passagem de informação, estando indisponível para tecnologias assistidas.

- Imagens de fundo: as imagens de fundo que contenham informações devem ser evitadas ou conter uma alternativa acessível adicional, já que não estão disponíveis em tecnologias assistidas.
- *Links*:
 - *Links* descritivos: O texto do *link* ou do item de navegação deve descrever exclusivamente seu destino ou função.
 - *Links* para formatos alternativos: *Links* para formatos alternativos devem indicar que uma página alternativa será aberta, caso contrário desorientará usuários com dificuldades cognitivas ou que se utilizam de tecnologia assistida.
 - Combinação de *links* repetidos: *Links* repetidos para o mesmo recurso devem ser combinados em um único *link*, o que auxiliará os usuários a navegar rapidamente pelo conteúdo, especialmente aqueles que dependem de tecnologia assistida.
- Notificações:
 - Notificações inclusivas: devem ser visíveis e audíveis.
 - Notificações do sistema operacional: devem ser utilizadas as notificações padrão do sistema operacional quando disponíveis e de forma apropriada.
 - Mensagens de erro e correção: devem ser claras.
 - *Feedback* e assistência: *Feedback* ou assistência não críticos devem ser fornecidos quando apropriado.
- *Scripts* e Conteúdos Dinâmicos:
 - Funcionamento progressivo: Aplicações e sites devem ser criados de forma a funcionar de maneira progressiva, garantindo uma experiência funcional para todos os usuários.
 - Controle de mídia: em apresentações de mídias devem existir botões para controles de pausa, parada ou mesmo ocultação dos controles.
 - Atualização de página: As atualizações automáticas de página não devem ser usadas sem prévio aviso, podendo impactar tecnologias assistidas, como por exemplo leitores de tela.
 - Tempos de espera: os tempos de resposta devem ser ajustáveis, algumas pessoas podem não ser capazes de responder dentro do tempo esperado.
 - Controle de entrada: Interações de entrada devem ser adaptáveis, de forma a permitir que usuários com deficiências motoras possam ajustar.
- Estrutura:

- Título único para páginas ou telas: Todas as páginas ou telas devem conter um único e identificável título.
- Cabeçalho: O conteúdo deve fornecer uma estrutura lógica e hierárquica de cabeçalho, de acordo com o que é suportado pela plataforma.
- Contêineres e marcadores: contêineres devem ser usados para descrever a estrutura da página ou tela, de acordo com o que é suportado pela plataforma.
- Grupos de elementos: controles, objetos e elementos de interface agrupados devem ser representados como um único componente acessível.
- Textos Equivalentes:
 - Alternativas para conteúdos não textuais: deve haver uma breve descrição da intenção ou propósito do conteúdo, imagem, objeto ou elemento.
 - Conteúdo decorativo: imagens decorativas devem ser escondidas de tecnologias assistidas.
 - Dicas e informações complementares: as dicas de ferramentas não devem repetir o texto do link ou outras alternativas.
 - Tarefas, marcas e propriedades: elementos devem conter propriedades de acessibilidade apropriados.
 - Formatação visual: não deve ser utilizado apenas a formatação visual para transmitir um determinado significado ou mensagem.

Conforme abordado em (SILVA, 2020), apesar das diretrizes de acessibilidade propostas pela BBC serem compreensíveis e de fácil interpretação, pode-se dizer que a W3C possui diretrizes não cobertas pela BBC, conforme apresentado abaixo:

- Montante de informações: em telas de aplicações móveis, é fundamental a redução da quantidade de informações apresentadas na tela se comparadas com as versões de *desktop*.
- Posição dos títulos: os formulários devem possuir seus títulos dispostos acima, ao invés de estarem ao lado da página.
- Teclado: todas as funcionalidades devem ser operáveis sem a necessidade da utilização de um teclado.
- Gestos: o seu emprego deve ser fácil, sem a necessidade de percorrer caminhos específicos.

- Posição dos elementos interativos: devem estar posicionados de forma que o usuário possa identificá-los facilmente.
- Orientação da tela: aplicações móveis devem suportar as duas orientações da tela, sendo possível de ser percebida facilmente por tecnologias assistidas.
- Posicionamento dos elementos: as informações importantes devem estar visíveis, sem que haja a necessidade de rolagem da tela para identificá-las.
- Instruções: é fundamental a inserção de títulos ou instruções que auxiliem o usuário a inserir as informações requeridas.
- Ajuda: devem estar disponíveis e de fácil identificação, mesmo com tecnologias assistidas.
- Facilidade na entrada de dados: pode haver a substituição de volumes de texto de entrada por menus de seleção, botões *radio*, caixas de seleção ou mesmo por entrada automática de dados, desde que estes estejam claramente informados ao usuário e contemplados por tecnologias assistidas.

Apesar do alto número de padrões e diretrizes definidos tanto pela BBC quanto pela W3C, é de conhecimento que estas definições não são completas, e que outros grupos ou companhias ou consórcios podem definir novos padrões de acordo com as necessidades de seus usuários.

2.2.3 eMAG - Modelo de Acessibilidade em Governo Eletrônico

Em 2004, o governo brasileiro iniciou os trabalhos relativos às definições dos padrões de acessibilidade e criou o Modelo de Acessibilidade Brasileiro, o eMAG (GOVERNMENT, 2014), com base em padrões internacionais como o WCAG. Este modelo foi proposto com o objetivo de ser um norteador para o desenvolvimento de portais Web acessíveis como forma de promover a inclusão social na população brasileira. Na data de 07 de maio de 2007, a Portaria nº 3 institucionalizou o padrão eMAG, tornando-o obrigatório no âmbito do governo federal ou de instituições públicas ou privadas que possuem relação com organizações federais. Posteriormente, em 2015, foi sancionada a lei Brasileira de nº 13.146, que estabelece as normas gerais de acessibilidade, incluindo as áreas de sistemas de informação e conteúdos digitais.

A versão 3.1 do eMAG, de Abril de 2014, define 45 recomendações de acessibilidade que estão divididas em seis seções. Diferentemente da WCAG, o eMAG não define prioridades ou níveis de conformidade para suas recomendações. A seguir é apresentada uma visão geral das recomendações deste padrão brasileiro:

- 1 Marcação: as 9 recomendações desta seção visam orientar os desenvolvedores a organizar as camadas do código, respeitando os padrões *web* de forma lógica e semântica. Além disso direciona a disponibilização de *links* diretos que facilitarão principalmente a navegação pelo usuário que se utiliza de tecnologias assistidas, bem como a não abertura de instâncias (abas ou janelas) sem a prévia solicitação do usuário.
- 2 Comportamento (DOM): as 7 recomendações descritas na seção visam orientar a utilização e controle da navegação. Independentemente do tipo de entrada utilizado (teclado, *mouse* ou mesmo *touchscreen*), os objetos programáveis devem ser acessíveis e as páginas não devem ser atualizadas ou redirecionadas automaticamente sem a autonomia do usuário. Nesta seção também se aborda o tema cintilação, que podem causar sérios danos a usuários sensíveis.
- 3 Conteúdo/Informação: as 12 recomendações desta seção visam garantir que o usuário possuirá as informações relevantes de cada elemento da tela, como por exemplo idioma da página ou conteúdo, textos descritivos e claros de objetos, imagens e links. Palavras incomuns, abreviaturas e siglas também devem conter suas descrições claramente. Por fim, direciona que estas informações estejam disponíveis em tecnologias assistidas.
- 4 Apresentação/Design: nesta seção estão as 4 recomendações relacionadas à disponibilização visual dos objetos da aplicação. São citados taxa de contraste mínimo, não utilização apenas da cor como elemento para transmissão de informações, redimensionar a tela sem perdas de funcionalidades e por fim possibilitar que o elemento com foco esteja destacado dos demais.
- 5 Multimídia: contemplando 5 recomendações, esta seção tem a função de direcionar que hajam alternativas e controles para o usuário na apresentação de informações de conteúdos multimídia. Legendas, audiodescrição, áudios alternativos, botões de controle são citados como fundamentais para o atendimento a esta categoria.
- 6 Formulário: as 8 recomendações visam disponibilizar formulários acessíveis, com textos descritivos, etiquetas e/ou orientações sobre cada item, manutenção da leitura

lógica através de tecnologia assistida, evitando alterações automáticas no contexto. Caso hajam erros nas inserções de dados, deverá haver textos explicativos e que sinalizem os locais a serem corrigidos. Por fim, estratégias de segurança também são abordadas nesta seção.

No detalhamento das recomendações mais complexas, são apresentados exemplos que auxiliam no entendimento e direcionamento dos desenvolvedores. Em decorrência do eMAG ter sido concebido baseado no WCAG, para cada recomendação existe o respectivo relacionamento com a WCAG.

2.3 *Trabalhos Relacionados*

O objetivo desta proposta de pesquisa é investigar se as avaliações dos aplicativos publicadas nas lojas oficiais estão relacionadas à acessibilidade da aplicação. Com exceção dos estudos iniciais realizados no início deste projeto (ELER; ORLANDIN; OLIVEIRA, 2019) e de uma pesquisa utilizando os dados disponibilizados por ele (TAMJEED, 2020) com a intenção de automatizar a identificação automática de avaliações relacionadas a acessibilidade, nenhum outro estudo foi encontrado com este mesmo objetivo. Portanto, nesta seção são mencionados os estudos publicados que fazem análise de outros aspectos relacionados às avaliações dos usuários publicadas nas lojas de aplicativos.

Com o advento das lojas de aplicativos (e.g. *Google Play Store* e *Apple Store*), a relação entre desenvolvedor e usuário começou a ser alterada. Avaliações de usuários passaram a ser postadas publicamente causando impactos tanto nas notas da aplicação (*ratings*) quanto no número de vezes que o software é baixado (*downloads*), e desta forma tornando-se para o desenvolvedor uma importante fonte de compreensão do seu público (Pagano; Maalej, 2013). De acordo com (FU *et al.*, 2013), baseado em mais de 13 milhões de avaliações, foi possível uma análise da exigência dos usuários de acordo com a categoria da aplicação, observando-se que os mesmos tendem a ser mais tolerantes para a categoria jogos do que para demais aplicações.

A oportunidade do cliente de expôr sua opinião com textos livres deu origem a um grande número de demandas (MCILROY *et al.*, 2016), o que traz aos desenvolvedores a dificuldade de identificar as necessidades a serem tratadas nas próximas entregas, especialmente para softwares populares com alta quantidade de opiniões postadas. Este

cenário tem promovido pesquisas sobre formas de sumarização (Iacob; Harrison, 2013; IACOB; HARRISON; FAILY, 2014; FU *et al.*, 2013) e interpretação destes textos, incluindo o emprego conjunto de diferentes técnicas para aprendizado de máquina (Panichella *et al.*, 2015).

Em (Palomba *et al.*, 2015), (PALOMBA *et al.*, 2018) e (LI; ZHANG; STEFANIDIS, 2018) foram realizadas pesquisas que associaram as avaliações às disponibilizações de versões do software (*releases*). As conclusões foram de que as notas aumentam quando há implementações que visam atender às opiniões dos usuários. Mesmo o pequeno volume de textos informativos trata-se de uma valiosa fonte de informações, que permite tanto a correção de erros de difícil identificação nos testes quanto a implantação de novas funcionalidades e recursos não-funcionais.

Embora diversos trabalhos tenham sido realizados para entender os tipos de demandas dos usuários e de que forma elas eram tratadas e utilizadas para o planejamento das novas versões dos aplicativos (Iacob; Harrison, 2013; Pagano; Maalej, 2013; IACOB; HARRISON; FAILY, 2014; MCILROY *et al.*, 2016; Di Sorbo *et al.*, 2017; Ciurumelea *et al.*, 2017; LI; ZHANG; STEFANIDIS, 2018; Pelloni *et al.*, 2018; Panichella *et al.*, 2015), nenhum estudo fez uma diferenciação ou um detalhamento sobre os aspectos de acessibilidade abordados pelos usuários.

A seguir são apresentados detalhes dos estudos relacionados a esta pesquisa.

2.3.1 *User Feedback in the AppStore: An Empirical Study*

O estudo de Pagano e Maalej (2013) já apresentava a importância das avaliações dos usuários na engenharia de requisitos. O estudo exploratório considerou mais de um milhão de avaliações, identificando que a maior parte era enviada logo após o lançamento da aplicação, com redução rápida ao longo do tempo. Os conteúdos dos textos causam impacto direto no número de *downloads*, sendo que as avaliações negativas referentes às deficiências da aplicação são tipicamente destrutivas e perdem detalhes de contexto e experiência do usuário.

Trata-se de um artigo de 2013, quando ainda estava sendo desvendado o relacionamento entre desenvolvedores, lojas de aplicativos e usuários que podem expôr suas avaliações abertamente, ou seja, formadores de opinião.

O artigo cita a observação de que o tamanho dos textos é maior para aplicações mais caras. Além disso, claramente os usuários tornam-se raivosos (enviando textos com insultos), principalmente quando houve compra da aplicação.

Em 2013, quando da elaboração do estudo, concluiu-se que as avaliações são uma ótima fonte para os desenvolvedores entenderem seu público, porém foi salientada a falta de ferramenta que permita a análise e agrupamento das avaliações que facilitem a mineração das informações pelos desenvolvedores.

2.3.2 *How can i improve my app? Classifying user reviews for software maintenance and evolution*

No trabalho de Panichella *et al.* (2015) é apresentado um estudo sobre as avaliações de aplicações móveis utilizando a combinação de três técnicas: Processamento de Linguagem Natural (PLN), Análise de Texto (AT) e Análise de Sentimento (AS). Os resultados sugerem ser possível a utilização destas informações para a evolução das aplicações pelos desenvolvedores.

As principais conclusões referentes às técnicas são que a combinação das mesmas produz melhor resultado na classificação e entendimento dos textos, e que a Análise de Sentimento sozinha não apresenta valores significativos, porém agregando valor quando combinada com as demais.

A metodologia do estudo consistiu em: (i) definir uma taxonomia para identificar os textos relacionados à manutenção e evolução do software; (ii) extração dos desejos dos usuários, considerando tanto novas funcionalidades quanto correções; (iii) aprendizado das técnicas e execução das mesmas combinadas; (iv) avaliação dos resultados. A taxonomia foi gerada manualmente por 2 autores do trabalho, e revisadas por um aluno de PhD, a partir de 300 e-mails produzidos em lista de discussão sobre 2 projetos de código fonte aberto, bem como se utilizando de técnica qualitativa publicada em 1967. Para a avaliação foi utilizada base de dados de trabalho anterior contendo as principais aplicações disponíveis nas lojas de aplicativos. Para a avaliação da metodologia foi feita a classificação manual por dois autores utilizando uma amostragem de 18% (1.421) das avaliações para um determinado software, que possuía 7.696 avaliações no total. Apenas 31 casos (2,81%) foram categorizados como "outros".

Os resultados sugerem que os usuários utilizam padrões para respostas quando a intenção é reportar um problema, porém quando a intenção é sugerir uma nova funcionalidade são percebidos diferentes padrões, o que dificulta a identificação. Além disso AS sozinha não apresenta valores significantes, no entanto agrega valor quando em combinação com outras técnicas.

Como ameaça de validade da base de estudo, pode-se citar a má interpretação do julgamento manual das classificações, porém mitigado com processo realizado em 2 etapas, por pessoas distintas, resultando em questionamentos para apenas 2,81% dos casos.

Como ameaças à validade interna citou-se inicialmente que os textos podem ser classificados em mais de uma intenção. Para isso eles analisaram o histórico das comunicações bem como ordenaram as classificações, permitindo alocar na de pior cenário. Outra ameaça à validade interna é o sobreajuste, cujo modelo se adequa muito bem ao oráculo, porém não necessariamente apresentará os mesmos resultados para o conjunto completo. Para reduzir esta ameaça foram executados os processos de aprendizagem sobre 20% das amostras, repetindo o processo 100 vezes e aplicando validação cruzada 10 vezes.

Como ameaças externas citou-se que o processo foi executado para um grupo específico de aplicações. Para reduzir esta ameaça foram selecionados casos de diferentes categorias, de duas lojas distintas.

2.3.3 *Analyzing reviews and code of mobile apps for better release planning*

No artigo de Ciurumelea *et al.* (2017) é apresentado o método *User Request Referencer* (URR) que permite ao desenvolvedor a obtenção mais rápida do código de seu software relacionado à avaliação do usuário. O estudo validou a possibilidade de categorizar a avaliação tanto em alto quanto baixo nível, salientando que o processo ainda necessitava de mais treinamentos. Foi estimada uma economia de tempo de até 75% para estas identificações do código, facilitando o processo para os casos em que as avaliações estão melhor explicadas.

O método URR se utiliza dos seguintes passos: (i) Definição da Taxonomia; (ii) Classificação das avaliações dos usuários; (iii) Localização do código fonte. A definição da taxonomia foi feita a partir de amostragens de 39 aplicações. Aqueles que possuíam

muitas avaliações a quantidade foi limitada a 200. O volume total de avaliações analisadas manualmente foi de 1.566.

Para a classificação de alto nível, tanto as avaliações positivas quanto as de classificações negativas são importantes para o desenvolvedor, o qual necessitará atuar mais fortemente ou poderá decidir atuar em outra funcionalidade. Para a classificação de baixo nível, a mesma é importante para o desenvolvedor identificar com precisão o ponto merecedor de atenção pelo desenvolvedor. A extração das funcionalidades foi feita utilizando Aprendizado de Máquina, de acordo com a taxonomia definida anteriormente. Foi decidido não considerar a avaliação em estrelas dada pelo usuário. Considerou-se a razão de 20% das avaliações como sendo base para aprendizagem.

Para a localização do código fonte, foi utilizado *Apache Lucene API*, de acordo com as etapas seguintes: (i) obtenção do código fonte de cada software; (ii) pré processamento de cada código fonte e avaliações considerando *stop word* em inglês e reduzindo distúrbios; (iii) indexação dos conteúdos do código fonte e avaliações; (iv) pré localização, feita através da criação de uma base de dados utilizando classificações das avaliações e localização das mesmas dentro do padrão de código fonte; (v) realização de pesquisa por relevância, onde neste teste foram obtidos resultados de 30%, com previsão de melhoras em trabalhos futuros.

As perguntas do artigo foram:

- RQ1: Até que ponto o Referenciador de Solicitações do Usuário organiza as avaliações de acordo com tarefas significativas de manutenção e evolução para desenvolvedores?
- RQ2: O URR recomenda corretamente os artefatos de software que precisam ser modificados para lidar com solicitações de usuários e reclamações?

Para este estudo foi utilizada uma base de 7.754 avaliações envolvendo 39 aplicações de código fonte abertos, que resultaram em 7.242 classes e 940.051 linhas de código Java. Dois experimentos foram realizados: O experimento I foi feito para responder à RQ1, em que se executou o URR com 20% dos textos e foi então solicitada a classificação manual dos mesmos, informando se o URR acertou ou não. Desta forma, não foi utilizado o mesmo conjunto de 1.566 avaliações utilizadas na elaboração da taxonomia. O experimento II foi feito para responder à RQ2, em que foi feito um oráculo com 91 avaliações.

As conclusões para a RQ1 é que é possível categorizar as avaliações tanto em alto quanto baixo nível, no entanto há a necessidade de melhorar o treinamento para alguns

casos. Foi estimada uma possível economia de tempo de até 75%, sendo as categorias mais importantes Uso, Recursos e Compatibilidade. Para RQ2, concluiu-se que a URR alcança resultados promissores, sendo que as melhores avaliações são mais fáceis de vincular.

2.3.4 *SURF: Summarizer of User Reviews Feedback*

Neste estudo de Di Sorbo *et al.* (2017) foi apresentada uma ferramenta denominada SURF (*Summarizer of User Reviews Feedback*) que se propõe analisar, classificar e sumarizar os textos das avaliações dos usuários com a intenção de facilitar a visualização pelos desenvolvedores. No estudo empírico participaram 12 desenvolvedores da Suíça, Itália e Holanda, que avaliaram a usabilidade da ferramenta. Um total de 12 aplicações e 2.622 textos de avaliações foram utilizados, de diferentes lojas e categorias de aplicativos.

A primeira etapa do processo foi a classificação da intenção dos usuários, utilizando Processamento em Linguagem Natural (PLN) e Análise de Sentimento. A segunda etapa classifica em diferentes agrupamentos, também se utilizando de através de PLN. A terceira etapa define então uma pontuação para cada avaliação, considerando vários itens como probabilidade da palavra estar em outras avaliações, número de palavras da avaliação e importância destas para a classificação. Apenas as avaliações com maiores pontuações são consideradas para a etapa seguinte, sendo que o corte foi de 70%. Por fim a etapa de geração da sumarização. A ferramenta gera um arquivo do tipo *XML* contendo as informações das avaliações, classificações e agrupamentos.

O artigo explica como utilizar a ferramenta, informa que foi disponibilizada em versão com interface para usuário de forma a facilitar tanto a inserção quanto a apresentação dos dados.

A avaliação foi realizada com 12 aplicações (subconjunto de um total de 17) e 2.622 avaliações, de diferentes lojas e categorias de aplicativos. Os resultados foram disponibilizados para pesquisadores, profissionais da indústria de software e desenvolvedores aleatoriamente, sendo que 9 participantes (de um total de 12) julgaram que os resultados auxiliam no entendimento dos textos. Foi estimada uma economia de tempo de 50% e assertividade de 92%.

As conclusões finais foram que futuramente a ferramenta poderia ser expandida para identificar a parte do código e evolução da extração dos dados.

2.3.5 BECLoMA: Augmenting stack traces with user review information

A ferramenta BECLoMA (*Bug Extractor, Classifier and Linker of Mobile Apps*) é apresentada no trabalho de Pelloni *et al.* (2018). Seu propósito é de vincular as avaliações de usuários aos processos de testes de aplicações, fornecendo aos desenvolvedores uma visão ampliada dos casos de testes e por consequência agilidade nas tomadas de ação. A ferramenta possui um *crawler* capaz de minerar avaliações do *Google Play Store* e já treinado para 6.600 avaliações. A parte central do software vincula os textos das avaliações aos códigos, retirando os métodos nativos e permitindo assim uma melhor visualização do problema.

A avaliação da ferramenta consistiu em responder à pergunta: Qual é a acurácia do BECLoMA em vincular as avaliações dos usuários aos relatórios de erro?

Foram realizados testes empíricos com 8 aplicações, número reduzido decorrente do esforço para geração de oráculos, construído por um desenvolvedor com 2 anos de experiência. A ferramenta apresentou precisão de 82%, *recall* de 75% e *F1-Score* de 78%, confirmando que o processo *Dice* para vinculação foi uma boa escolha, no entanto foi importante salientar que aproximadamente 18% dos casos que deveriam ser vinculados a ferramenta não conseguiu associar.

A conclusão do estudo cita que a ferramenta pode no futuro ser evoluída e ter uma grande usabilidade para os desenvolvedores identificar erros e depurar códigos, e como próximo passo foi citada a utilização da ferramenta em conjunto com a IDE *Android Studio*.

2.3.6 Why People Hate Your App — Making Sense of User Feedback in a Mobile App Store

A ferramenta WisCom é apresentada em (FU *et al.*, 2013) como sendo uma alternativa para sumarizações de avaliações de usuários, podendo auxiliar tanto grandes empresas de software como os pequenos desenvolvedores. No estudo foram utilizados mais de 13 milhões de avaliações da *Google Play Store*, em mais de 171 mil aplicações.

No estudo a ferramenta identificou inconsistências entre avaliações e classificações de usuários, os principais motivos pelos quais os usuários não gostam de uma aplicação, mudanças ao longo do tempo das avaliações e, por fim, tendências globais no mercado.

Permitindo a realização de *drill down* de avaliações em 3 níveis distintos, através da ferramenta é possível identificar casos de inconsistências e razões pelo qual o usuário gosta ou não da aplicação, e fornecer visões intuitivas da loja de aplicativos que permitam identificar preferências dos usuários com relação a diferentes aplicações.

A primeira parte do artigo apresenta a comparação dos textos das avaliações com as respectivas notas, sinalizando os casos discrepantes. A segunda parte do artigo tem a intenção de sinalizar os casos discrepantes para o operador da loja de aplicativos, desenvolvedores e até mesmo o próprio usuário.

Como o estudo envolveu mais de 13MM de avaliações, foi possível realizar uma análise independente da aplicação. O estudo conclui também que os usuários tendem a ser mais tolerantes para jogos do que outros tipos de aplicações, considerando softwares pagos.

2.3.7 *Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews*

A possibilidade e importância da rotulação das avaliações é apresentada no trabalho de Mcilroy *et al.* (2016). O estudo rotulou (em 11 classificações distintas) mais de 600 mil textos de 12 mil aplicações. O artigo cita que com as avaliações os desenvolvedores podem entender melhor as preocupações dos usuários, os proprietários de lojas de aplicativos podem identificar softwares anômalos e os usuários podem comparar aplicações semelhantes e decidir quais utilizar. Foi abordado que a natureza não estruturada e informal das avaliações, se utilizando muitas vezes de gírias e abreviações, complica a rotulagem automática de tais avaliações.

Além disso, o artigo cita que algumas empresas se especializaram em prover dados estatísticos e comparativos detalhados sobre avaliações para seus clientes, porém muitas vezes não é um trabalho orientado ao software.

As perguntas respondidas pelo estudo foram:

- RQ1 - Quantas avaliações contém múltiplos tipos de problemas apontados? A resposta foi que 22% (Apple) e 30% (Android) das avaliações são *multi-label*, tendo uma média de palavras um pouco maior (passando de 32 para 41).
- RQ2 - Quão bem é possível categorizar as avaliações em mais de um tipo de categoria? No estudo, a precisão das análises foram de até 66%, com recall de até 65%, sendo que a proposta de estudo futuro é de melhorar o processo de predição.

- RQ3 - A abordagem de multi categorias é funcional para os *stakeholders*? O artigo informa que sim, e para isso considera provas de conceito.

2.3.8 *Retrieving and Analyzing Mobile Apps Feature*

No artigo de Iacob e Harrison (2013) é proposto um processo denominado MARA (*Mobile App Review Analyzer*) para auxiliar na identificação de avaliações que apresentem solicitações de funcionalidades. No processo, a ferramenta se utiliza de *Latent Dirichlet Allocation* (modelo estatístico de Processamento de Linguagem Natural) para as informações.

Na investigação foi considerado que textos com sarcasmos, não estruturados, reduzidos, com erros de pontuação são algumas barreiras a serem ultrapassadas no entendimento das avaliações.

As conclusões do estudo mostram que 23% das avaliações sugerem novas funcionalidades ou alteração em casos existentes, e que a maioria das solicitações dos usuários diz respeito ao suporte aprimorado das aplicações, atualizações mais frequentes, novos níveis para jogos e mais opções de personalização.

2.3.9 *Online Reviews as First Class Artifacts in Mobile App Development*

No estudo Iacob, Harrison e Faily (2014) é apresentado o impacto das avaliações dos usuários no processo de engenharia de software, com argumentos explicando que os modelos e técnicas clássicas podem não ser os mais adequados. Baseado em evidências, o protótipo MARA visa oferecer integração das avaliações com o processo de engenharia de software, extraindo as solicitações de recursos e relatórios de erros.

Este novo estudo da ferramenta MARA estendeu a análise para classificação de erros, com aprendizagem baseada em exemplo de 3.279 avaliações randômicas, tratadas manualmente.

2.3.10 *User Reviews Matter! Tracking Crowdsourced - Reviews to Support Evolution of Successful Apps*

Uma nova abordagem, denominada CRISTAL, é apresentada por Palomba *et al.* (2015) com o propósito de rastrear as avaliações comprovando que desenvolvedores que implementam as solicitações dos usuários são recompensados em termos de notas. O processo se utiliza de mecanismos *crowdsourcing* sobre as avaliações das lojas de aplicativos, tendo como conjunto de estudo um total 100 aplicações Android.

Neste estudo foi também avaliada a possibilidade de relacionamento dos *commits* e *issues* com as avaliações dos usuários, porém apenas em um conjunto restrito de 10 aplicações, obtendo indicador *F-Measure* de 75% (77% de *precision* e 73% de *recall*).

As questões abordadas no estudo foram:

- RQa: Quão preciso é o CRISTAL na identificação de links entre avaliações informativas e *issues* e *commits*?

O artigo explica a possibilidade de associação entre as avaliações dos usuários e os respectivos *commits* das aplicações, sendo esta associação com uma precisão de 77% e recall de 73%.

- RQb: Até que ponto os desenvolvedores cumprem as avaliações ao trabalhar em um novo lançamento de aplicação?

Os resultados alcançados mostraram que, na média, os desenvolvedores implementam 49% das solicitações nas avaliações enquanto trabalham no lançamento de nova aplicação.

- RQc: Qual é o efeito de um mecanismo análise massiva de avaliações (para planejar e implementar mudanças futuras) no sucesso da aplicação?

Confirmado através de uma correlação de 0,59 entre a cobertura das avaliações e a alteração de nota entre a versão anterior e nova da aplicação. Este percentual foi validado informado que a análise qualitativa suporta parte dos resultados quantitativos encontrados.

Como ameaça do estudo citou-se que a nota da aplicação foi melhorada decorrente de outras formas, como por exemplo uma importante entrega (*feature*) disponibilizada. No entanto, a intenção do artigo é de apresentar uma correlação quantitativa ao invés de uma relação direta de causa e efeito. Também foram encontradas evidências através

de mineração dos dados e discussões entre os autores de casos em que o mesmo usuário melhorou a nota após a liberação de uma nova versão da aplicação.

A conclusão do artigo é que as notas das aplicações aumentam quando os desenvolvedores implementam alterações que visam responder às avaliações dos usuários.

2.3.11 *Crowdsourcing user reviews to support the evolution of mobile apps*

No trabalho de Palomba *et al.* (2018), a abordagem CRISTAL foi complementada com uma pesquisa junto a 73 desenvolvedores.

Os resultados qualitativos da pesquisa corroboram as informações quantitativas encontradas através da abordagem. Os desenvolvedores concordam que as avaliações representam uma valiosa informação, levando a correções difíceis de serem identificadas nos testes, implementação de recursos de sucesso e requisitos não funcionais. A grande maioria dos entrevistados (90%) acredita que estas implementações têm um efeito de sucesso da aplicação.

2.3.12 *Mobile App Evolution Analysis based on User Reviews*

O artigo de Li, Zhang e Stefanidis (2018) associa avaliações com versões disponibilizadas da aplicação (*releases*), porém tendo apenas como objeto de estudo a aplicação *Whatspapp*.

Ao longo do estudo determinou-se quais foram as principais versões da aplicação e então realizada uma análise de sentimento sobre as avaliações enviadas entre estas versões.

Após agrupamento das palavras chave, verificou-se o comportamento destes grupos ao longo dos períodos. A conclusão foi de que se trata de um estudo investigativo exploratório, propondo que estudos futuros abordem novas categorias de software, plataforma, e principalmente que um sistema possa ser desenvolvido com base neste método de forma a minerar as avaliações e prover dados contínuos aos desenvolvedores.

O processo consistiu em determinar quais foram as principais versões da aplicação e em seguida realizar uma análise de sentimento sobre as avaliações enviadas entre as publicações destas versões.

2.3.13 *Do Android App Users Care about Accessibility? An Analysis of User Reviews on the Google Play Store*

O estudo de Eler, Orlandin e Oliveira (2019) analisa as avaliações feitas pelos usuários de um conjunto de 700 aplicações para identificar os aspectos de acessibilidade mencionados. Os detalhes deste artigo são apresentados na Seção 4 pois foi o estudo de viabilidade que subsidiou esta proposta de pesquisa.

2.3.14 *Accessibility in User Reviews for Mobile Apps: An Automated Detection Approach*

O estudo de Tamjeed (2020) utilizou os dados da pesquisa de Eler, Orlandin e Oliveira (2019) que estão disponíveis publicamente² para propor e avaliar um mecanismo de aprendizado de máquina para identificar automaticamente avaliações de acessibilidade. Entretanto, os autores não analisaram as avaliações de outros aplicativos ou avançaram nas discussões sobre o tema.

2.4 *Considerações Finais*

Em decorrência de relevante parcela da população mundial possuir algum grau de deficiência, bem como do avanço tecnológico que permitiu um aumento considerável da utilização de software para dispositivos móveis, entende-se como fundamental a evolução das aplicações no que diz respeito à sua acessibilidade.

Apesar desta importância e mesmo em aplicações móveis populares, identifica-se problemas de acessibilidade conforme os padrões internacionalmente reconhecidos, o que leva ao indício de que não há a devida significância para este tema junto aos desenvolvedores, ou mesmo não há uma pressão de mercado que induza a esta priorização durante a elaboração de novas versões.

Considerando que as avaliações são uma importante fonte de retroalimentação para os desenvolvedores e dada a importância da acessibilidade no cenário mundial, entende-se que existe a necessidade de um estudo do relacionamento entre as avaliações em lojas de aplicativos, os aspectos de acessibilidade mencionados, e as informações cadastradas pelos desenvolvedores para as solicitações de modificações e alterações de fato realizadas. Nos

² <https://github.com/marceloeler/data-ihc2019>

estudos encontrados na literatura sobre análise de avaliações de usuários e evolução dos aplicativos, o tema acessibilidade não é explorado, apesar de alguns estudos classificar as avaliações dentro da categoria de interface com o usuário.

3 Proposta de Pesquisa

O objetivo geral deste trabalho de mestrado é descobrir se questões relacionadas à acessibilidade são tratadas pelos usuários de aplicações móveis em avaliações submetidas às lojas de aplicativos, e se tais avaliações são levadas em consideração pelos desenvolvedores e organizações responsáveis pela manutenção e evolução destas aplicações.

Para atender a este objetivo, foi adotada uma pesquisa exploratória envolvendo artigos relacionados a avaliações de usuários em lojas de aplicativos bem como levantamentos destas avaliações em conjunto com dados de solicitações de modificações e alterações em repositórios de códigos-fonte abertos (GitHub), visto não ter sido identificado estudo relacionando estas três bases de informações no que diz respeito ao tema acessibilidade.

Este capítulo apresenta as decisões metodológicas deste trabalho, tais como as questões de pesquisa, as atividades previstas e as configurações dos estudos definidos para atingir os objetivos propostos neste trabalho.

3.1 Questões de Pesquisa

Algumas questões de pesquisa gerais e específicas foram identificadas para guiar a execução deste trabalho:

- RQ1 - Os usuários de aplicações móveis mencionam aspectos relacionados à acessibilidade em suas avaliações publicadas nas lojas de aplicativos?
 - RQ1.1 - Quantas avaliações de usuários são relacionadas à acessibilidade e qual é a sua distribuição entre as aplicações avaliadas?
 - RQ1.2 - O quão diverso são os tópicos de acessibilidades mencionados nas avaliações dos usuários?
 - RQ1.3 - Qual é a relação entre as questões de acessibilidade abordadas pelos usuários e as notas da aplicação recebidas em cada avaliação?
- RQ2 - Os problemas de acessibilidade relatados nas avaliações dos usuários têm impacto na evolução das aplicações móveis?
 - RQ2.1 - Quantas solicitações de modificações (melhorias, correções de defeitos ou novas funções) registradas nos repositórios de código das aplicações estão relacionadas à acessibilidade?

- RQ2.2 - Quantas alterações realizadas nos códigos das aplicações estão relacionadas à acessibilidade?
- RQ2.3 - Qual é a correlação entre as avaliações relacionadas à acessibilidade das aplicações e as sugestões de modificações e alterações efetuadas nos códigos das aplicações.

3.2 *Atividades Planejadas e Decisões de Projeto*

A seguir estão listadas as atividades propostas para atingir os objetivos deste projeto de pesquisa e responder as questões levantadas neste trabalho.

3.2.1 Revisão Bibliográfica

Inicialmente, será realizada uma etapa de estudos envolvendo os principais conceitos de acessibilidade. Em particular, serão estudados padrões de acessibilidade internacionais como o WCAG (W3C, 2018b), o padrão da BBC (BBC, 2017) e o e-MAG (GOVERNMENT, 2014). Adicionalmente, serão analisados os estudos que utilizam avaliações de usuários no planejamento da evolução de aplicações móveis.

3.2.2 Seleção de aplicativos móveis

É necessário selecionar um conjunto de aplicações para a realização da investigação proposta. Alguns critérios foram definidos para esta seleção: as aplicações selecionadas devem estar publicadas em uma loja de aplicativos; as avaliações das aplicações realizadas pelos usuários devem ser públicas e possíveis de serem acessadas; e as aplicações devem ter código-aberto e devem estar disponíveis em repositórios de acesso público.

Considerando os critérios de seleção, decidiu-se eleger aplicações da plataforma Android por três razões principais: esta plataforma possui o maior mercado de aplicações móveis do mundo¹, com 72,92%; a maioria dos estudos sobre evolução de aplicações móveis são relacionados a esta plataforma; e o acesso aos dados da loja de aplicativos oficial (*Google Play Store*) é gratuito e existem ferramentas disponíveis para extrair as informações necessárias para a realização deste projeto.

¹ <http://gs.statcounter.com/os-market-share/mobile/worldwide> em 04/11/2020

Além disso, optou-se por analisar as aplicações indexadas no FDroid², um repositório que mantém um catálogo de aplicações gratuitas e de código fonte aberto. Este repositório possui mais de 3168 aplicações indexadas³ das mais diversas categorias, tamanhos, complexidades e popularidade, cujas informações podem ser acessadas por meio de um arquivo XML fornecido pela própria plataforma⁴.

Infelizmente, nem todas as aplicações indexadas nesta plataforma estão publicadas na *Google Play Store*. Portanto, a seleção das aplicações deve incluir apenas aquelas que estão publicadas na loja oficial do Android para que seja possível obter as avaliações dos usuários. Ainda, para analisar as solicitações de alterações e alterações realizadas nas aplicações é preciso que o código-fonte esteja disponível publicamente. Portanto, decidiu-se selecionar apenas aplicações cujos códigos-fonte estejam armazenados no repositório GitHub, pois este repositório oferece uma API (*Application Programming Interface*) para a consulta de informações sobre solicitações de modificações e alterações realizadas nas aplicações. No GitHub, as solicitações de alterações são denominadas de *issues* e as alterações realizadas no código são chamadas de *commit*.

3.2.3 Extração das avaliações dos usuários

Esta atividade consiste em obter as avaliações publicadas pelos usuários referentes às aplicações selecionadas. As avaliações consistem em uma nota atribuída para o aplicativo, um título para a avaliação (opcional) e um comentário escrito pelo usuário (opcional). As avaliações também podem conter comentários ou respostas para os usuários escritas pelos desenvolvedores ou pelas organizações responsáveis pela publicação do aplicativo na loja oficial.

A *Google Play Store* não oferece uma API pela qual se pode extrair os detalhes das avaliações dos usuários, portanto decidiu-se utilizar uma API não oficial chamada *google-play-api*⁵. Esta API foi construída com base em uma biblioteca que percorre automaticamente as páginas (*crawler*) da loja de aplicativos e transforma as informações coletadas em dados que podem ser consumidos por meio de uma API.

² <https://www.f-droid.org/>

³ Dados extraídos em 23/10/2020

⁴ <https://f-droid.org/repo/index.xml>

⁵ <https://github.com/facundoolano/google-play-api>

Para extrair as avaliações da loja de aplicativos, é preciso construir um software que utilize a API *google-play-api*, a fim de coletar as informações necessárias das aplicações móveis selecionadas para este estudo. Para cada avaliação, as seguintes informações são fornecidas: um identificador, o nome do usuário que fez a avaliação, a data da avaliação, a nota que o usuário atribuiu ao aplicativo, o título do comentário, o corpo do texto do comentário, a data de resposta dada ao usuário (se houver), e o texto da resposta dada o usuário (se houver).

3.2.4 Extração das solicitações de modificações e alterações de código

Esta atividade consiste em obter as solicitações de modificações e as alterações realizadas no código das aplicações selecionadas para este estudo. No GitHub, as solicitações de modificações (*issues*) podem ser criadas pelos próprios desenvolvedores ou qualquer outro usuário da plataforma. As alterações (*commits*) são realizadas pelos desenvolvedores/colaboradores do projeto e podem conter diversas informações, tais como rótulos, descrição das alterações efetuadas e as modificações efetuadas no código (inserção, deleção, etc).

O GitHub fornece uma API⁶ para os desenvolvedores utilizarem para extrair informações dos repositórios hospedados na plataforma. As informações das solicitações de modificações (*issues*) obtidas por meio desta API são as seguintes: um identificador, a data de criação e a situação da solicitação (aberta, atendida), o título da solicitação, o corpo do texto da solicitação, um rótulo para a solicitação (defeito, nova funcionalidade, melhoria, etc), dados sobre o usuário que fez a solicitação, comentários feitos por outros usuários sobre a solicitação, entre outras. As informações das alterações (*commits*) obtidas por meio desta API são as seguintes: um identificador, o autor da alteração, um texto explicando a alteração realizada, e ponteiros para informações detalhadas sobre as modificações realizadas no código.

Para obter as informações necessárias para este estudo, é preciso construir um software capaz de utilizar a API fornecida pelo GitHub para coletar e organizar as informações das solicitações de modificações e alterações realizadas no código das aplicações.

⁶ <https://developer.github.com/v3/>

3.2.5 Seleção de avaliações, solicitações de modificações e alterações

Uma das atividades mais importantes deste projeto de pesquisa é a identificação das avaliações, das solicitações de modificações e das alterações realizadas, que estão relacionadas à acessibilidade das aplicações selecionadas. Para a realização desta atividade, decidiu-se fazer a seleção com base em duas etapas: fazer um filtro dos itens analisados por meio de palavras-chave; e realizar análise manual para eliminar os falso-positivos. O conjunto de palavras-chave será criado com base na análise das diretrizes de acessibilidade de padrões nacionais e internacionais de acessibilidade. A análise manual será feita pelo autor desta proposta, pelo orientador, e por outros colaboradores do mesmo grupo de pesquisa.

3.2.6 Análise das avaliações, solicitações de modificações e alterações

Após a seleção por meio de palavras-chave e análise manual, as avaliações dos usuários, as solicitações de modificações e as alterações realizadas nos aplicativos serão analisadas para responder as questões de pesquisa definidas na Seção 3.1.

3.3 *Contribuições Esperadas*

As principais contribuições esperadas desta pesquisa são as evidências de que não só as aplicações móveis são pouco acessíveis, mas que também os requisitos relacionados à acessibilidade digital raramente são abordados por usuários em suas avaliações publicadas nas lojas de aplicativos, e consequentemente esses requisitos não são abordados durante a evolução das aplicações móveis. O entendimento de que existe uma lacuna nas demandas feitas pelos usuários para que as aplicações se tornem mais acessíveis pode ser útil para a criação de ações cujo objetivo seja fornecer aos usuários com alguma deficiência informações e recursos suficientes para utilizar esta ferramenta tão útil para direcionar a evolução e a melhoria da qualidade das aplicações.

Além da contribuição geral desta proposta de pesquisa, pode-se citar as seguintes contribuições pontuais:

- Disponibilização de um conjunto de dados referentes às avaliações, solicitações de modificações e alterações, para posteriores estudos relacionados;
- Criação de um conjunto de palavras-chave relacionadas à acessibilidade para a identificação de textos que tratem de aspectos mencionados em padrões como o WCAG e a BBC
- Código fonte em Python que permita a extração dos dados utilizados neste trabalho: avaliações do *Google Play Store*, solicitações de modificações e alterações cadastradas no Github.

3.4 Cronograma

A Tabela 4 apresenta as atividades e o cronograma previstos para a realização deste projeto de pesquisa. Algumas atividades apresentadas referem-se ao projeto de pesquisa em si e são relacionadas às tarefas descritas na Seção 3.2, e algumas atividades referem-se a etapas do curso de mestrado. A atividade denominada *Estudo Piloto* refere-se a um estudo exploratório realizado para investigar a viabilidade do projeto proposto. Os detalhes desse estudo estão apresentados no próximo capítulo desta proposta.

Tabela 4 – Cronograma

| Atividades | 2019 | | | 2020 | | | 2021 | |
|-----------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| | Fev-Abr | Mai-Ago | Set-Dez | Jan-Abr | Mai-Ago | Set-Dez | Jan-Abr | Mai-Ago |
| Créditos em disciplinas | ■ | | | | | | | |
| Revisão Bibliográfica | ■ | ■ | ■ | ■ | | ■ | | |
| Seleção de aplicativos | | ■ | | | | ■ | | |
| Extração de avaliações | | ■ | | | | ■ | | |
| Estudo exploratório | ■ | ■ | ■ | | | | | |
| Extração de <i>issues/commits</i> | | | | | | ■ | | |
| Definição de palavras-chave | | ■ | ■ | | | ■ | | |
| Seleção dos itens extraídos | | | | | | ■ | | |
| Análise manual | | | | | | | ■ | |
| Análise dos dados | | | | | | | ■ | ■ |
| Exame de qualificação | | | | | | ■ | | |
| Escrita da dissertação | | | | | | | ■ | ■ |
| Depósito da dissertação | | | | | | | | ■ |
| Publicação de artigos | | | ■ | | | | | ■ |

4 Estudo exploratório e atividades já realizadas

Este capítulo apresenta um estudo exploratório realizado com o objetivo de verificar a viabilidade e aperfeiçoar a proposta com base nas descobertas e desafios enfrentados durante o processo. Destaca-se que este estudo teve o foco apenas na análise das avaliações de acessibilidade e não explorou as solicitações de modificações e as alterações realizadas no código das aplicações.

Todas as decisões de projeto e metodologia adotadas estão em conformidade com o que foi descrito no Capítulo 3. Além disso, também são descritas as atividades já realizadas do cronograma proposto.

4.1 Questões de pesquisa

As seguintes questões de pesquisa foram definidas para este estudo exploratório:

- RQ1 - Quantas avaliações de usuário são relacionadas à acessibilidade e qual é a sua distribuição?
- RQ2 - Qual é a diversidade de tópicos de acessibilidade abordados nas avaliações dos usuários?
- RQ3 - Quais são as notas associadas às avaliações que abordam aspectos de acessibilidade?

4.2 Seleção de aplicações móveis

Os critérios de seleção das aplicações para este estudo seguem os mesmos definidos na Seção 3.2.5. Portanto, foram selecionadas aplicações da plataforma Android que satisfaziam as seguintes condições: estavam disponíveis na *Google Play Store* e o código-fonte se encontrava armazenado em um repositório público da plataforma GitHub. Ao todo, apenas 701 aplicações dentre as mais de 2.000 indexadas no FDroid¹ satisfizeram esses critérios.

A Tabela 5 mostra a distribuição de alguns atributos das aplicações selecionados por meio de estatística descritiva utilizando o resumo dos cinco números (mínimo amostral,

¹ Dados de julho/2019

quartil inferior, mediana, quartil superior, máximo amostral), a média e o desvio padrão. A maioria das aplicações tem até 9 atividades, mas há exemplos muito maiores, como o “Slide for Reddit”, por exemplo, que possui 91 atividades.

Tabela 5 – Estatísticas da amostra das avaliações do *Google Play Store*

| | Atividades | Nota | Avaliações | Instalações | Comentários |
|------------------|------------|------|------------|-------------|-------------|
| Mínimo | 1 | 0 | 0 | 0 | 0 |
| Quartil Inferior | 2 | 4 | 50 | 1K | 4 |
| Mediana | 5 | 4,3 | 130 | 10K | 22 |
| Quartil Superior | 9 | 4,6 | 836 | 50K | 145 |
| Máximo | 92 | 5 | 3.6M | 100M | 4.480 |
| Média | 8,16 | 4,16 | 12.4K | 550K+ | 305,4 |
| Desvio Padrão | 10,5 | 0,79 | 151,6K | 5,7M | 840,8 |

A maioria das aplicações possui uma boa nota e foi avaliada por até 836 usuários. O número de avaliações apresentado na tabela é diferente do número de comentários, pois quando um usuário faz uma avaliação ele necessariamente atribui uma nota, mas não é obrigado a inserir um comentário. Existem aplicações que possuem um número muito alto de avaliações, como o Telegram, que possui 3,6 milhões de avaliações. O Telegram também é a aplicação mais instalada (100 milhões de instalações). Por outro lado, o número de avaliações não é muito alto, visto que a maioria das aplicações recebeu no máximo 145 comentários em suas avaliações. O número máximo de avaliações desta amostra é de 4.480 em razão das limitações da API utilizada. Esta limitação não teve muito impacto neste estudo exploratório porque apenas 15 aplicações (2%) possuem mais de 4.480 avaliações registradas na *Google Play Store*. No momento do estudo piloto, ficou definido que para as novas coletas de avaliações seriam consideradas outras alternativas para que não houvesse limite no número de avaliações retornadas pela API.

4.3 Extração e seleção das avaliações

Uma aplicação escrita na linguagem Python foi desenvolvida para consumir os dados da *Google Play Store* utilizando a API *google-play-api*². Como mencionado anteriormente, na versão utilizada esta API só permitia o retorno de no máximo 4.480 avaliações de cada aplicação. Além disso, só foram recuperadas avaliações escritas em inglês, que é o idioma padrão definido pela API.

² <https://github.com/facundoolano/google-play-api>

A seleção das avaliações que possivelmente estão relacionadas a algum aspecto de acessibilidade da aplicação foi feita por meio da utilização de palavras-chave aplicadas aos comentários dos usuários. Para isso, um conjunto de palavras-chave foi definido com base nas diretrizes de acessibilidade da BBC (BBC, 2017), o qual possui 54 recomendações que são classificadas em onze diferentes categorias: áudio e vídeo (5); design (12); editorial (3); foco (6); formulários (6); imagens (2); links (3); notificações (4); scripts e conteúdo dinâmico (4); estrutura (4); e texto equivalente (5). Este padrão foi selecionado porque tem o foco em aplicações móveis e possui exemplos de implementação e de teste para cada recomendação, o que permite compreender melhor os tipos de problemas de acessibilidade tratados no guia.

Foram definidas 213 palavras-chave com base na análise de cada recomendação do guia. A Tabela 6 mostra exemplos das palavras-chave utilizadas. Note que foram consideradas variantes das palavras (exemplo: *cannot see* e *can't see*, ou *color* e *colour*, ou *impaired* e *impairment*) para garantir que nenhuma avaliação relevante fosse excluída. Infelizmente, não é possível capturar casos em que as palavras foram escritas com a grafia errada. A lista completa das palavras-chave pode ser vista em um repositório criado para armazenamento dos dados deste estudo³.

Tabela 6 – Exemplos de palavras-chave utilizadas para selecionar avaliações relacionadas à acessibilidade das aplicações

| Categorias | Palavras-chave |
|-------------------|--|
| Gerais | accessibility, disability, screen reader, Talkback, operable, impaired, impairment |
| Áudio/Vídeo | subtitle, sign language, audio description, transcript, autoplay, mute, volume |
| Design | contrast, background color, blind, flicker, visual cue, touch size, overlap, font size, dark/light mode, eyestrain, seizure, can't see |
| Editorial | consist. label, language, visual/audio cue |
| Foco | focusable, control focus, keyboard trap, focus order, navigable |
| Formulários | unique label, missing label, content description, input type, input format, focusable |
| Imagens | image of text, hidden text, text alternative, background image |
| Links | link description, unique desc., duplicate link, alternative format |
| Notificações | inclusive, haptic, vibration, feedback, alert dialog, understandable, unfamiliar |
| Conteúdo dinâmico | animated content, page refresh, automatic refresh, timeout, adaptable, input sign |
| Estrutura | page title, screen title, heading, header |
| Texto equivalente | alternative text, non-visual, blind, screen reader, content description |

³ <https://github.com/marceloeler/data-ihc2019>

O uso de palavras-chave pode trazer muitos falsos positivos, uma vez que várias palavras utilizadas podem ter conotações diferentes de acordo com o contexto. Por isso foi realizada uma análise manual de todas as avaliações que possuíam pelo menos uma palavra-chave definida. Neste estudo, apenas uma pessoa fez a análise manual das avaliações. Além disso, as avaliações não selecionadas por palavras-chave não foram analisadas para saber se alguma delas não foi selecionada pela ausência de alguma palavra-chave relevante e que não foi utilizada. Durante a análise manual, as avaliações confirmadas também foram classificadas em: requisições, quando o usuário reporta algum problema de acessibilidade ou solicita alguma modificação ou adição para tornar a aplicação mais acessível; ou elogios, quando o usuário parabeniza a aplicação pelo seu nível de acessibilidade.

4.4 Análise dos resultados

As avaliações selecionadas com base nas palavras-chave e na validação manual foram analisadas para responder as questões de pesquisa propostas no estudo exploratório. A seguir as respostas às questões de pesquisa são apresentadas em detalhes.

RQ1 - Quantas avaliações de usuário são relacionadas à acessibilidade e qual é a sua distribuição?

No total, foram extraídas 214.053 avaliações completas (com nota e comentário) das 701 aplicações da amostra. Dessas avaliações, apenas 5.076 foram pré-selecionadas por meio das palavras-chave. Após a análise manual, o número de avaliações relacionadas à acessibilidade das aplicações foi reduzido para 2.663, o que representa apenas 1,24% de todas as avaliações completas extraídas.

As avaliações de acessibilidade não estão igualmente distribuídas entre as aplicações da amostra. Só 40% dos casos (276) possuem pelo menos uma avaliação relacionada à acessibilidade. É importante ressaltar que 92% (197.419) de todas as avaliações extraídas foram realizadas para essas 276 aplicações, indicando que quanto menos avaliações, menor são as chances de haver uma avaliação de acessibilidade. De fato, a correlação de Spearman calculada entre o número de avaliações e o número de avaliações de acessibilidade é de 0,73, o que representa uma forte correlação.

As avaliações de acessibilidade não estão distribuídas igualmente entre as 276 aplicações. Quase 65% das avaliações de acessibilidade (1.745) estão concentradas em

apenas 35 aplicações, e dentre essas, três possuem mais de 100 avaliações de acessibilidade enquanto as demais têm entre 19 e 79 avaliações cada. A Figura 1 mostra a distribuição de avaliações de acessibilidade entre as aplicações da amostra, com exceção das 35 aplicações mencionadas anteriormente. É possível perceber que a grande maioria das aplicações têm menos do que nove avaliações de acessibilidade, enquanto cerca de metade da amostra tem até três avaliações do tipo.

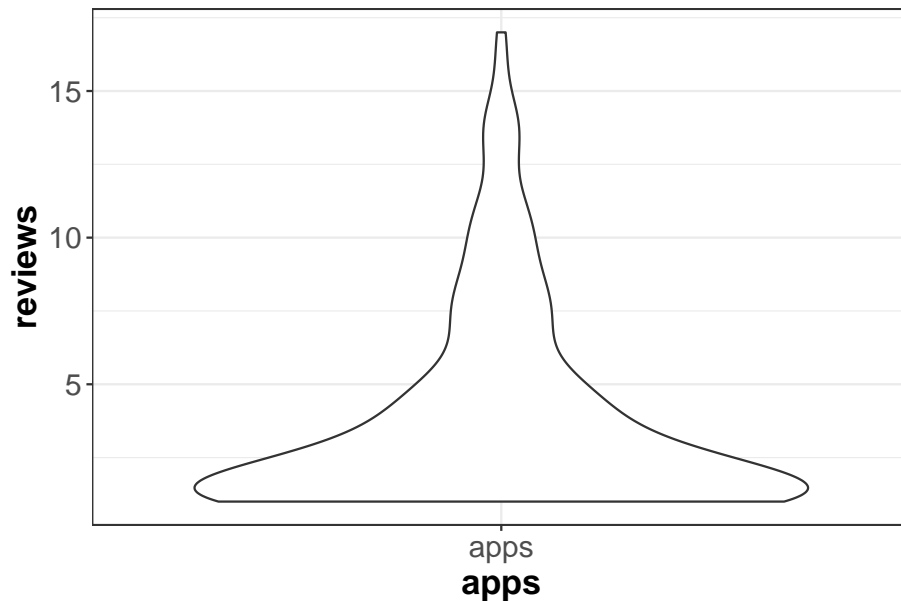


Figura 1 – Distribuição das avaliações de acessibilidade entre as aplicações da amostra

A Figura 2 mostra a distribuição da proporção entre as avaliações de acessibilidade e o número total de avaliações completas de cada aplicação. No geral, as avaliações de acessibilidade representam menos de 4% de todas as avaliações feitas para cada aplicação, e para metade das aplicações elas representam menos de 1,5%. Para uma pequena proporção de aplicações, as avaliações de acessibilidade representam entre 4% e 9% do total de avaliações recebidas. Existem algumas aplicações para as quais essa proporção varia entre 10% e 16%. Há alguns casos em que o número de avaliações completas é tão pequeno que uma simples avaliação de acessibilidade representa uma grande proporção, como é o caso das aplicações MuPDF mini, Booky McBookface, e GameDealz, cujas proporções são de 50% ($1/2$), 40% ($2/5$) e 33% ($1/3$), respectivamente.

RQ2 - Qual é a diversidade de tópicos de acessibilidade abordados nas avaliações dos usuários?

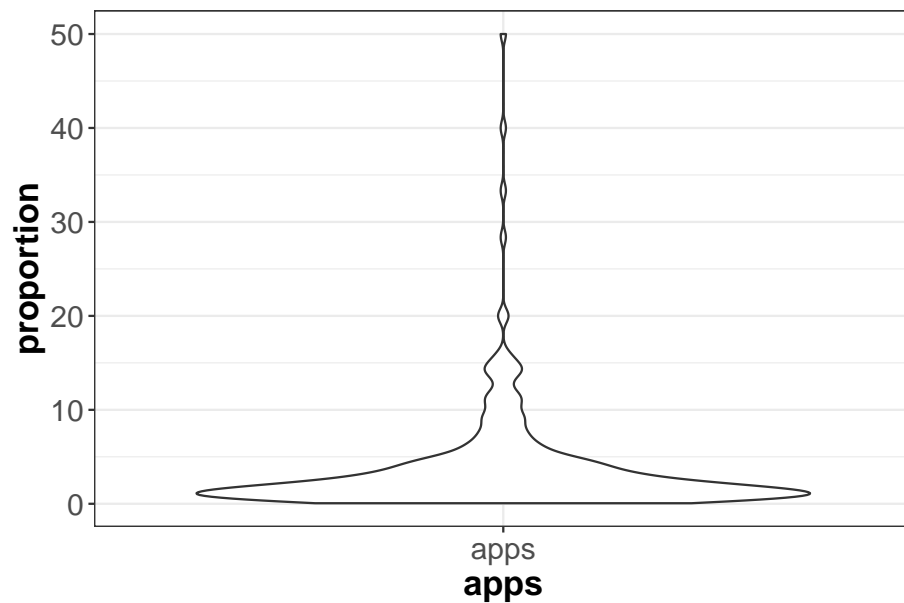


Figura 2 – Distribuição da proporção entre avaliações de acessibilidade e o total de avaliações de cada aplicação

O objetivo desta questão de pesquisa é entender se os usuários abordam diferentes problemas de acessibilidade em suas avaliações ou se a maioria das avaliações trata-se de um conjunto restrito de problemas. A Tabela 7 mostra o número de avaliações de acessibilidade e aplicações associadas a cada palavra-chave utilizada para selecionar as avaliações. Nesta tabela só estão apresentadas as palavras-chave para as quais pelo menos uma avaliação estava associada.

Claramente, as avaliações de acessibilidade estão concentradas em um pequeno conjunto de palavras-chave. Seis das palavras-chave mais populares são mencionadas em mais de 100 avaliações de diferentes aplicações. Por exemplo, a palavra-chave *dark mode* aparece em 628 avaliações de 128 aplicações distintas, o que representa quase metade da amostra de aplicações.

Como várias palavras-chave podem estar relacionadas a uma mesma diretriz ou princípio de acessibilidade (exemplo: *dark mode*, *night mode* e *black mode*), as palavras-chave e as respectivas quantidades de avaliações foram agrupadas por tópicos (diretrizes, princípios ou temas) específicos de acessibilidade para facilitar a análise da diversidade de características abordadas.

A Tabela 8 mostra, para cada tópico de acessibilidade (Coluna 1), o número de avaliações distintas (Coluna 2), a proporção das avaliações em relação às avaliações de acessibilidade (Coluna 3), o número de aplicações distintas que tiveram uma avaliação

Tabela 7 – Número de avaliações e aplicações por palavra-chave

| palavra | avals. | apps | palavra | avals. | apps | palavra | avals. | apps |
|----------------|--------|------|------------------|--------|------|--------------------|--------|------|
| dark mode | 645 | 128 | metadata | 20 | 12 | grouped | 3 | 3 |
| zoom | 491 | 80 | too bright | 20 | 8 | seizures | 3 | 1 |
| customization | 309 | 50 | haptic | 16 | 10 | select language | 3 | 3 |
| font size | 214 | 74 | scaling | 16 | 11 | understandable | 3 | 3 |
| volume | 146 | 40 | control key | 15 | 5 | vibration feedback | 3 | 3 |
| cannot see | 128 | 57 | voice command | 14 | 10 | actionable | 2 | 1 |
| accessibility | 74 | 44 | text-to-speech | 13 | 9 | audio cue | 2 | 2 |
| readable | 72 | 47 | eyestrain | 12 | 9 | missing label | 2 | 2 |
| change font | 68 | 44 | strain | 12 | 9 | navigable | 2 | 2 |
| hard to see | 58 | 41 | backgrd. image | 11 | 8 | verbose | 2 | 2 |
| backgrd. color | 50 | 30 | screen reader | 11 | 10 | captcha | 2 | 2 |
| light mode | 42 | 27 | change language | 10 | 8 | audio description | 1 | 1 |
| mute | 42 | 25 | small widget | 10 | 9 | container | 1 | 1 |
| contrast | 40 | 31 | stop button | 10 | 5 | distinguishable | 1 | 1 |
| subtitle | 40 | 7 | impaired | 9 | 9 | input type | 1 | 1 |
| adjustable | 34 | 23 | text reflow | 9 | 3 | keyboard language | 1 | 1 |
| blind | 31 | 24 | timeout | 9 | 7 | page refresh | 1 | 1 |
| header | 31 | 22 | consistency | 7 | 7 | page title | 1 | 1 |
| overlap | 31 | 25 | epilepsy | 7 | 1 | sign language | 1 | 1 |
| pause button | 27 | 17 | assistance | 6 | 5 | svg image | 1 | 1 |
| flicker | 26 | 19 | colour coding | 5 | 5 | switch device | 1 | 1 |
| spacing | 26 | 17 | transcript | 5 | 5 | touch target | 1 | 1 |
| migraine | 25 | 3 | default language | 4 | 4 | adjust size | 1 | 1 |
| input method | 23 | 11 | older device | 4 | 3 | adjust colour | 1 | 1 |
| autoplay | 21 | 17 | visual cue | 4 | 4 | | | |

relacionada (Coluna 4), e a proporção de aplicações que possuem pelo menos uma avaliação (Coluna 5). As diretrizes gerais da BBC não foram utilizadas para a organização dos tópicos apresentados na tabela porque eles são muito gerais e podem englobar muitas palavras-chave ao mesmo tempo.

O tópico *Theme/Mode* é o mais popular entre as avaliações uma vez que 27% das avaliações de acessibilidade estão relacionadas a este tópico, e quase metade das aplicações tem este tipo de avaliação. Este tópico agrega todas as avaliações que mencionam modos de cores e temas das interfaces das aplicações. O segundo tópico mais popular é o relacionado a funcionalidades de *zoom* (19% das avaliações de acessibilidade e 30% das aplicações). O terceiro tópico mais popular está relacionado à personalização da aplicação, representando 13% das avaliações e relacionados a 23% das aplicações. Diversos outros tópicos estão presentes em um número menor de avaliações de acessibilidade, mas muitos deles estão relacionados com um grande número de aplicações.

Tabela 8 – Distribuição de avaliações de acessibilidade e aplicações por tópico de acessibilidade

| Tópico | Avals. | % (Avals.) | Apps | % (Apps) |
|--------------------|--------|------------|------|----------|
| Theme/Mode | 726 | 27% | 144 | 52% |
| Zoom | 506 | 19,0% | 83 | 30,1% |
| Customization | 351 | 13% | 63 | 23% |
| Media | 309 | 11,6% | 82 | 29,7% |
| Font | 249 | 9,4% | 83 | 30,1% |
| Contrast | 218 | 8% | 91 | 33% |
| Impairment | 135 | 5,1% | 70 | 25,4% |
| Flickering | 87 | 3,3% | 31 | 11,2% |
| Accessibility | 74 | 2,8% | 44 | 15,9% |
| Size | 66 | 2,5% | 45 | 16,3% |
| Input alternatives | 54 | 2,0% | 25 | 9,1% |
| Feedback | 19 | 0,7% | 11 | 4,0% |
| Language | 15 | 0,6% | 11 | 4,0% |

Além da análise de todas as avaliações em conjunto, também foram realizadas análises das diversidades de tópicos de acessibilidade de cada aplicação isoladamente. Figura 3 mostra a distribuição de palavras-chave encontradas nas avaliações de acessibilidade de cada aplicação (lado esquerdo). No geral, há pouca diversidade de palavras-chave nas avaliações de cada aplicação analisada individualmente. A correlação de Spearman entre a diversidade de palavras-chave e o número de avaliações em cada aplicação é de 0,93, o que indica que a diversidade aumenta conforme aumenta o número de avaliações.

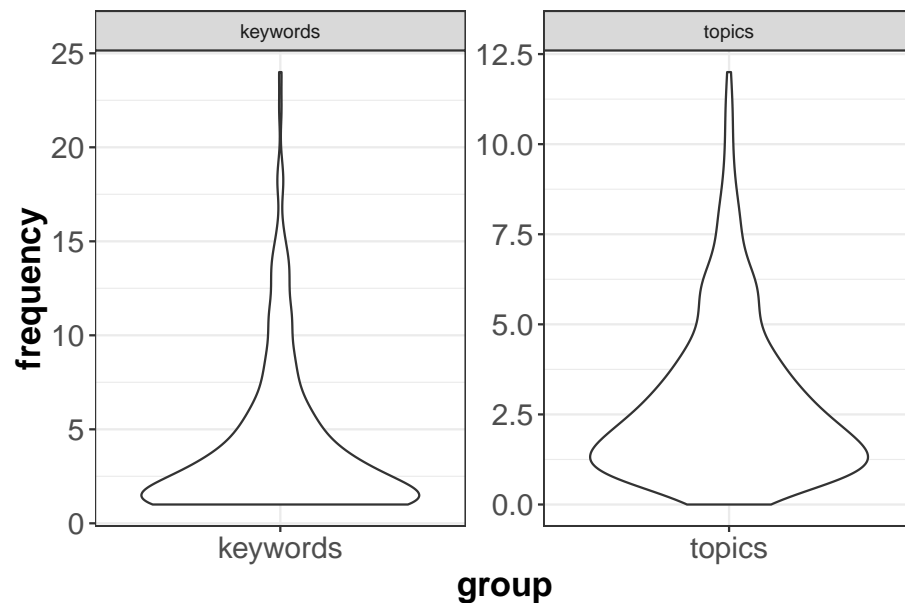


Figura 3 – Distribuição de tópicos de acessibilidade e palavras-chave nas avaliações de cada aplicação

Para a maioria das aplicações, as avaliações de acessibilidade estão relacionadas a no máximo seis palavras-chave. Para metade das aplicações, no máximo três palavras-chave são mencionadas. Poucas aplicações têm avaliações que mencionam mais de 15 palavras-chave. Uma das aplicações que foge ao padrão é o *Cool Reader*, pois em suas avaliações foram encontradas 36 palavras-chave distintas.

A diversidade de palavras-chave não é uma indicação de que diversos aspectos de acessibilidade são tratados uma vez que muitas delas estão relacionadas a um mesmo problema, e por isso uma análise por tópico deve ser realizada. A Figura 3 mostra a distribuição de tópicos de acessibilidade relacionados a cada avaliação de acessibilidade (lado direito). Note que as avaliações de acessibilidade para a maioria das aplicações estão concentradas entre dois e quatro tópicos, enquanto algumas aplicações possuem avaliações associadas a quase todos os tópicos abordados. A aplicação *Calculator*, por exemplo, tem avaliações de acessibilidade que abordam 12 tópicos de acessibilidade distintos. Embora a correlação de Spearman entre a diversidade de tópicos e o número de avaliações seja de 0,87, o que representa uma correlação forte, esta aplicação possui apenas 60 avaliações.

RQ3 - Quais são as notas associadas às avaliações que abordam aspectos de acessibilidade?

Durante a validação manual (ver Seção 4.3), as avaliações foram classificadas entre requisições (76%) ou elogios (24%). As 645 avaliações que elogiam a acessibilidade estão distribuídas por 110 aplicações, enquanto as requisições estão distribuídas por 262 aplicações. Um dos propósitos desta questão de pesquisa é entender se os usuários dão uma nota boa para a aplicação quando fazem um elogio e uma nota ruim quando fazem alguma requisição. A Figura 4 mostra a comparação das notas recebidas pelas avaliações classificadas nessas duas categorias. Aparentemente, não há nenhuma diferença significativa entre as notas atribuídas aos dois grupos.

Adicionalmente, deseja-se descobrir se avaliações associadas a diferentes palavras-chave ou tópicos de acessibilidade possuem notas maiores ou menores. A Figura 5 mostra as notas associadas a cada tópico de acessibilidade (ver Tabela 8). Neste caso, somente foram consideradas avaliações que contém requisições. Para alguns tópicos específicos, mesmo que haja requisição dos usuários para tornar as aplicações mais acessíveis, as notas são, em sua maioria, 4 ou 5 (exemplo: *customization*, *font*, *media*, *size* e *theme*). Para

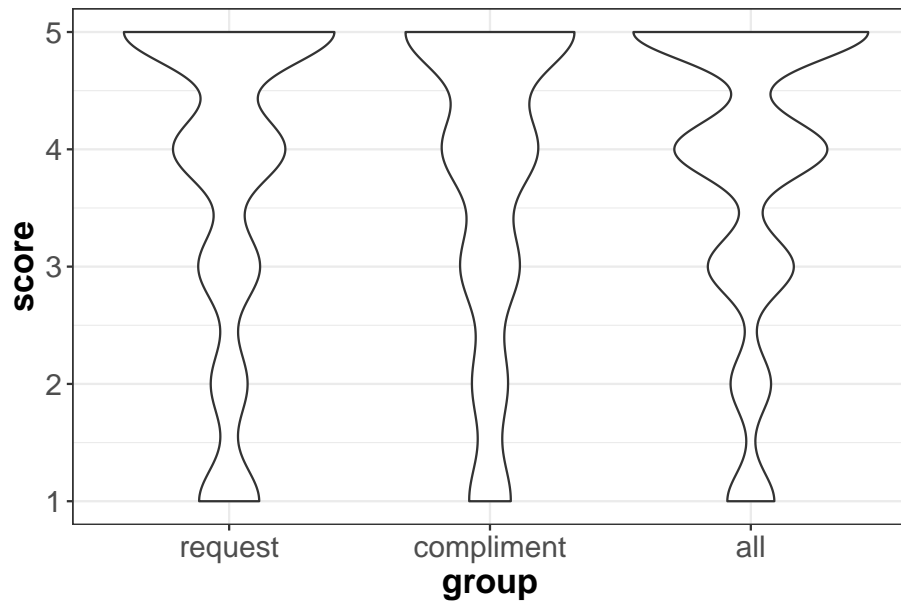


Figura 4 – Comparação entre as notas associadas a avaliações que contém requisições com as avaliações que contém elogios à acessibilidade das aplicações

outros tópicos, entretanto, as notas estão concentradas em valores mais baixos, de 2 a 4 (exemplo: *contrast*, *flickering* e *zoom*).

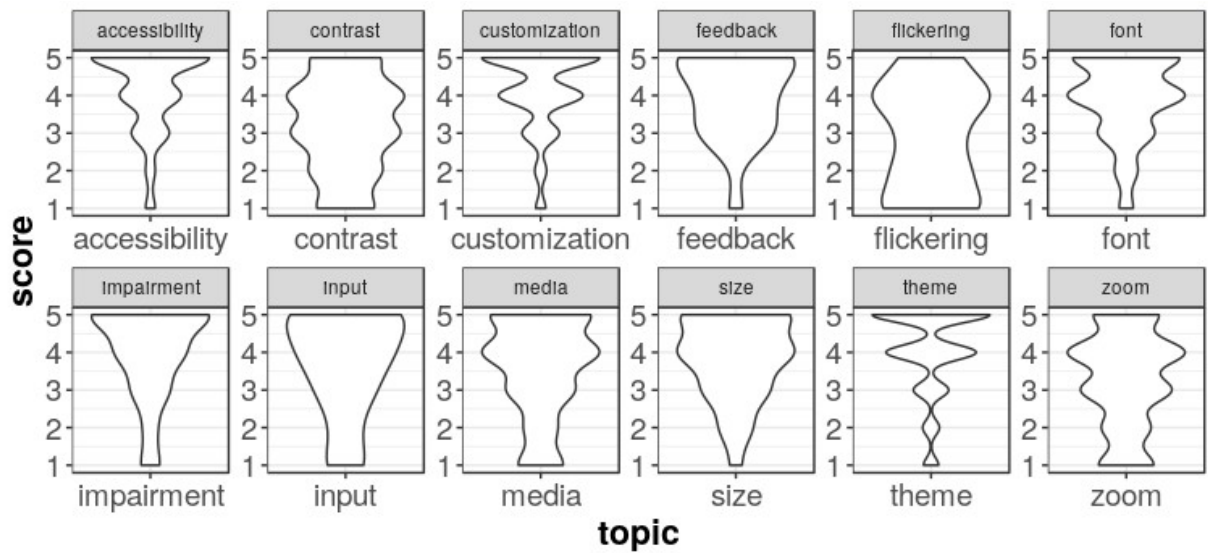


Figura 5 – Notas associadas às avaliações que abordam um tópico de acessibilidade (cf. Tabela 8)

A Figura 6 mostra as notas associadas a avaliações que contém algumas das palavras-chave utilizadas neste estudo. Para algumas palavras-chave as notas estão concentradas nos valores máximos (*accessibility*, *blind*, *change font*, *dark mode*, *customization*, *haptic*,

etc), para outras palavras-chave as notas são significativamente mais baixas (*cannot see*, *epilepsy*, *flicker*, *migraine*, *readable*, *text reflow* e *too bright*).

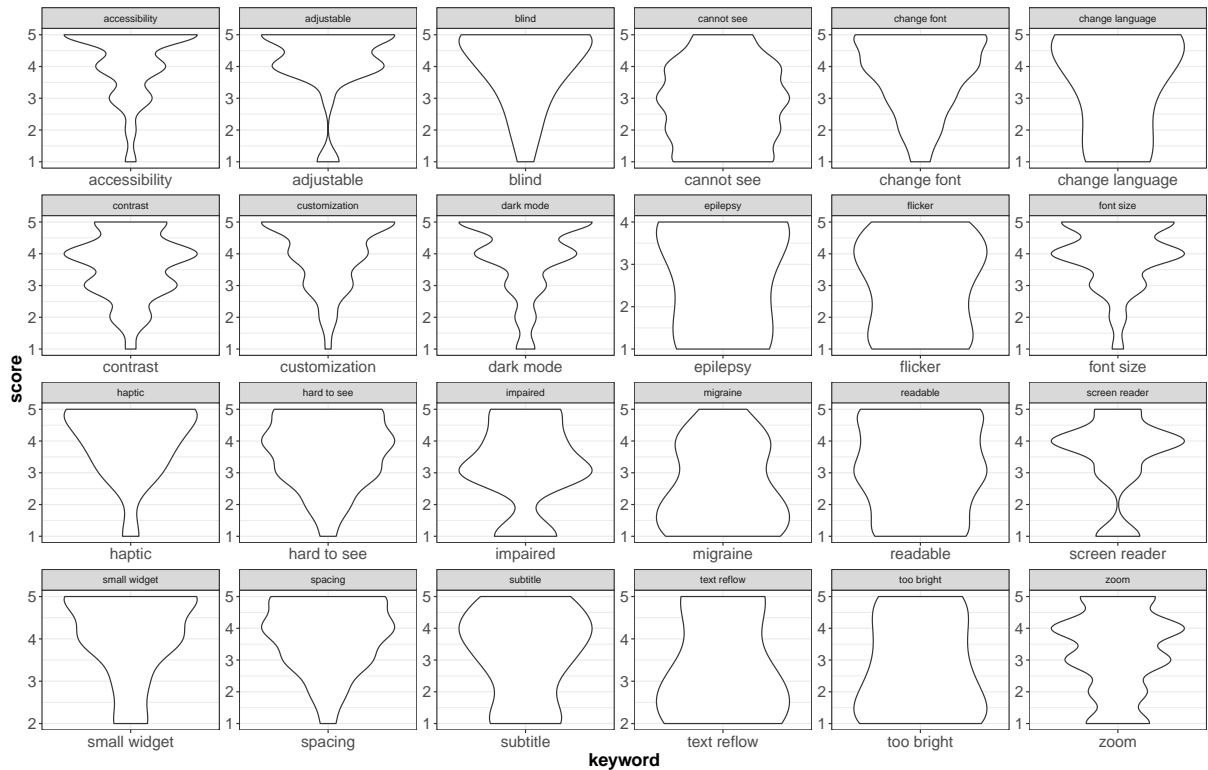


Figura 6 – Notas associadas aos comentários que apresentam uma determinada palavra-chave

Aparentemente, os usuários só atribuem notas baixas às aplicações quando encontram problemas de acessibilidade que são barreiras para a utilização da aplicação. Alguns exemplos de comentários de usuários encontrados nas avaliações que contém palavras-chave associadas às menores notas estão apresentados a seguir:

- “*Very little contrast. Bad for old eyes.*”
- “*Wish it also handled older open doc formats - and that it would reflow text on zoom*”
- “*The flashing backgrounds could trigger an epilepsy and migraine*”
- “*It’s ok but the editing icons come up in white over the image.. which obviously means if you have a white image you cannot see them. Seriously - how could a developer get something so obvious that wrong?*”
- “*Widget Text color is black. Unreadable*”

- *“Too bright now on this update now. Can you option to inverted color? (black instead of white). Going to have to uninstall. Sorry.”*

4.5 Discussão

A maioria das avaliações de acessibilidade encontradas foram feitas por usuários que não necessariamente estavam tratando diretamente da acessibilidade da aplicação. Ainda que essas avaliações mencionem aspectos da aplicação que influenciam a sua acessibilidade, nem todas elas são requisições relacionadas a barreiras encontradas pelos usuários, mas solicitações para melhorar sua experiência de uso. Embora poucos usuários se identifiquem como pessoas com deficiência em suas avaliações, percebe-se que as requisições realizadas por eles são associadas a barreiras reais encontradas durante o uso da aplicação, acompanhadas de notas mais baixas.

Em geral, pessoas com deficiência tendem a ser enfáticas sobre as barreiras que as impedem de usar plenamente a aplicação. A seguir são apresentados alguns exemplos de comentários feitos por pessoas com deficiência:

- *“I am legally blind and use Voice Assistant. This app does not interface well with Voice Assistant. It jumps to reading the logo - when trying to get to next group or text to be read. Extremely frustrating! Please fix right away!”*
- *“The layout of the app is miserable for those with visual impairments. Please rethink your layout for those who can’t read the tiny clues (even on a tablet!)”*
- *“Some buttons are unlabelled (I’ve figured out which ones are repeat and shuffle - I still don’t think I know what some are). Also the sleep timer is half-way accessible - when I was using the slider to select the minutes it wouldn’t read the amount of time I selected (might be the fault of the screen reader though - I’m using the Mobile Accessibility screen reader from Code Factory).”*
- *“Hello. If this app become accessible to Talkback (android screen reader) I wish the best for you developers. Actually we blinds can use telegram only on ios because it is accessible with Voiceover (IOS screen reader). One of my friends needs it because he does not have money to buy an apple product but he has a cheap android device and he needs telegram on it because of their university channel on telegram. If this app becomes accessible I will rate it 5 or more!”*

- “*What are you trying to do? Give someone who has epilepsy a seizure? I almost did. Thanks a lot, idiots*”

A escassez e a homogeneidade das avaliações de acessibilidade associadas às boas notas recebidas em geral pelas aplicações parece indicar que ou os usuários com deficiência não utilizam essas aplicações ou eles tendem a não realizar avaliações. Nos estudos ampliados que serão realizados com base neste estudo exploratório, pretende-se fazer análise separadas de avaliações que claramente foram escritas por pessoas com alguma deficiência ou que enfrentaram uma barreira real para a utilização da aplicação pela falta de acessibilidade.

Por fim, considerando que as avaliações dos usuários são uma poderosa ferramenta para guiar a evolução de aplicações móveis (IACOB; HARRISON; FAILY, 2014; Panichella *et al.*, 2015; PALOMBA *et al.*, 2018), esses resultados preliminares sugerem que isso não tem sido utilizado apropriadamente para pressionar desenvolvedores e organizações na melhoria da acessibilidade das aplicações.

4.6 Publicações

Os resultados deste estudo exploratório foram divulgados em artigo completo publicado na trilha principal do XVIII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC 2019) (ELER; ORLANDIN; OLIVEIRA, 2019). Destaca-se que este artigo recebeu o prêmio de melhor artigo do IHC 2019.

4.7 Novas configurações de estudo e atividades realizadas

O planejamento e os resultados do estudo apresentado anteriormente foram importantes para validar a existência de avaliações que tratam da acessibilidade das aplicações, ainda que de forma escassa, e o processo adotado para seleção e análise dos comentários escritos pelos usuários. Para a continuação deste trabalho de mestrado, pretende-se fazer um estudo estendido e aperfeiçoado para obter resultados mais robustos. A seguir estão listadas as diferenças do estudo anterior e dos estudos que serão executados em seguida:

- As aplicações foram selecionadas com base no estado dos repositórios em junho de 2019. Uma nova análise será realizada para atualizar a lista de aplicações.

- Coleta de avaliações da *Google Play Store*: serão coletadas as avaliações atualizadas de todas as aplicações selecionadas.
- Quantidade de avaliações coletadas: no estudo exploratório apresentado havia o limite de extração de 4.480 avaliações por aplicação, mas as atualizações realizadas na API utilizada permitem agora extrair todas as avaliações.
- Identificação de avaliações de acessibilidade: serão utilizadas palavras-chave e avaliação manual. Entretanto, diferentemente do estudo exploratório, as palavras-chave serão definidas com base na análise do WCAG 2.1 e não mais com base nas diretrizes da BBC. Essa decisão foi tomada porque o WCAG é um padrão utilizado internacionalmente e é utilizado como referência em muitas leis que promovem o desenvolvimento de software e conteúdo digital acessível.
- Coleta de solicitações de modificações e alterações realizadas nos repositórios de código e cruzamento de informações com as avaliações coletadas e avaliadas.

Algumas dessas atividades já foram realizadas e apresenta-se os resultados como segue:

- **Amostra de aplicações:** foi realizada uma seleção de aplicações com base nos critérios definidos na Seção 3.2.2: as aplicações precisam estar indexadas no repositório FDroid, estarem disponíveis na *Google Play Store* e ter código-fonte armazenado em repositório público no GitHub. Além disso, as aplicações selecionadas devem ter pelo menos uma avaliação completa (nota e comentário) registrada na *Google Play Store*. Ao todo, 921 aplicações foram selecionadas para este estudo⁴.
- **Extração de avaliações:** foram extraídas as avaliações completas (nota e comentário) das 921 aplicações da amostra. Ao todo, foram extraídas 1,5 milhão de avaliações. Apesar do grande número de avaliações, mais de 80% das aplicações possui menos do que 1.000 avaliações, e poucas aplicações possuem mais do que 100 mil.
- **Extração de solicitações de modificações (*issues*) e alterações (*commits*):** os repositórios de código das aplicações da amostra foram analisados e destes extraídas as solicitações de modificações e alterações realizadas no código. Ao todo, foram extraídas aproximadamente 41 mil solicitações de modificações e 1,1 milhão de alterações.

⁴ Com base em dados coletados em 23 de outubro de 2020

5 Considerações finais

As demandas realizadas por usuários em suas avaliações publicadas em lojas de aplicativos impactam as ações dos desenvolvedores e das empresas na correção de defeitos e no planejamento de novas funcionalidades que serão disponibilizadas em novas versões. Esta proposta de pesquisa visa investigar um cenário até o momento não explorado por outros estudos, que são os aspectos de acessibilidade abordados nas avaliações e correlação deste tipo de demanda com as solicitações de modificações e alterações de fato realizadas nas aplicações móveis para que sejam mais acessíveis.

Até o momento, foi realizado um estudo piloto com aproximadamente 700 aplicações e que forneceram evidências suficientes para motivar um estudo mais amplo e detalhado. Para este novo estudo, já foram selecionados 900 aplicações e extraídas 1,5 milhão de avaliações, 41 mil solicitações de mudanças e mais de 1 milhão de alterações realizadas no código das respectivas aplicações investigadas.

Os resultados desta pesquisa vão mostrar se os usuários que possuem alguma deficiência ou que encontram barreiras de acessibilidade estão utilizando esta importante ferramenta que são as avaliações publicadas nas lojas de aplicativos para fazer suas demandas por melhorias da acessibilidade das aplicações. O estudo inicial realizado mostrou que as demandas deste tipo são mínimas, mas um estudo ampliado com mais avaliações coletada e com novas palavras-chave para seleção de aplicativos será importante para dar mais robustez aos resultados.

A confirmação de que as avaliações são pouco exploradas para demandar melhorias na acessibilidade das aplicações pode ser utilizada como subsídio de ações cujo objetivo seja promover o maior engajamento de usuários com deficiência na exigência de aplicativos com maior qualidade e com mais acessibilidade uma vez que os comentários dos usuários tem tido impacto nas ações dos desenvolvedores no que se refere à evolução de suas aplicações móveis.

Referências¹

ACOSTA-VARGAS, P.; SALVADOR-ULLAURI, L.; JADÁN-GUERRERO, J.; GUEVARA, C.; SANCHEZ-GORDON, S.; CALLE-JIMENEZ, T.; LARA-ALVAREZ, P.; MEDINA, A.; NUNES, I. L. Accessibility assessment in mobile applications for android. In: NUNES, I. L. (Ed.). *Advances in Human Factors and Systems Interaction*. Cham: Springer International Publishing, 2020. p. 279–288. ISBN 978-3-030-20040-4. Citado na página 8.

ALSHAYBAN, A.; AHMED, I.; MALEK, S. Accessibility issues in android apps: State of affairs, sentiments, and ways forward technical. In: *International Conference on Software Engineering*. [S.l.]: ACM, 2020. p. 1–12. Citado na página 8.

BARZILAI-NAHON, K.; BENBASAT, I.; LOU, N. Factors influencing users' intentions to make the web accessible to people with disabilities. In: IEEE. *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*. [S.l.], 2008. p. 126–126. Citado na página 8.

BBC. *The BBC Standards and Guidelines for Mobile Accessibility*. 2017. Disponível em: <https://www.bbc.co.uk/guidelines/futuremedia/accessibility/mobile>. Citado 4 vezes nas páginas 13, 14, 37 e 44.

BBC. *BBC - Homepage*. 2020. Disponível em: <https://www.bbc.com/>. Nenhuma citação no texto.

Ciurumelea, A.; Schaufelbühl, A.; Panichella, S.; Gall, H. C. Analyzing reviews and code of mobile apps for better release planning. In: *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. [S.l.: s.n.], 2017. p. 91–102. Citado 4 vezes nas páginas 8, 9, 24 e 26.

Di Sorbo, A.; Panichella, S.; Alexandru, C. V.; Visaggio, C. A.; Canfora, G. Surf: Summarizer of user reviews feedback. In: *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. [S.l.: s.n.], 2017. p. 55–58. Citado 3 vezes nas páginas 9, 24 e 28.

ELER, M. M.; ORLANDIN, L.; OLIVEIRA, A. D. A. Do android app users care about accessibility? an analysis of user reviews on the google play store. In: *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems (IHC '19)*. [S.l.: s.n.], 2019. Citado 3 vezes nas páginas 23, 34 e 54.

Eler, M. M.; Rojas, J. M.; Ge, Y.; Fraser, G. Automated accessibility testing of mobile apps. In: *2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST)*. [S.l.: s.n.], 2018. p. 116–126. Citado na página 8.

FREIRE, A. P.; RUSSO, C. M.; FORTES, R. P. M. A survey on the accessibility awareness of people involved in web development projects in brazil. In: *Proceedings of the 2008 International Cross-disciplinary Conference on Web Accessibility (W4A)*. New York, NY, USA: ACM, 2008. (W4A '08), p. 87–96. ISBN 978-1-60558-153-8. Disponível em: <http://doi.acm.org/10.1145/1368044.1368064>. Citado na página 8.

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

FU, B.; LIN, J.; LI, L.; FALOUTSOS, C.; HONG, J.; SADEH, N. Why people hate your app: Making sense of user feedback in a mobile app store. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2013. (KDD '13), p. 1276–1284. ISBN 978-1-4503-2174-7. Disponível em: <http://doi.acm.org/10.1145/2487575.2488202>. Citado 3 vezes nas páginas 23, 24 e 29.

GAY, G.; MILLER, F.; WEST, C. *Digital Accessibility as a Business Practice – Essential Skills for Business Leaders*. Canada: Independent, 2018. Citado na página 7.

GOVERNMENT, B. F. *eMAG - Modelo de Acessibilidade em Governo Eletrônico*. 2014. Disponível em: <http://emag.governoeletronico.gov.br/>. Citado 3 vezes nas páginas 13, 21 e 37.

Iacob, C.; Harrison, R. Retrieving and analyzing mobile apps feature requests from online reviews. In: *2013 10th Working Conference on Mining Software Repositories (MSR)*. [S.l.: s.n.], 2013. p. 41–44. ISSN 2160-1860. Citado 3 vezes nas páginas 9, 24 e 31.

IACOB, C.; HARRISON, R.; FAITY, S. Online reviews as first class artifacts in mobile app development. In: MEMMI, G.; BLANKE, U. (Ed.). *Mobile Computing, Applications, and Services*. Cham: Springer International Publishing, 2014. p. 47–53. ISBN 978-3-319-05452-0. Citado 4 vezes nas páginas 9, 24, 31 e 54.

INAL, Y.; RIZVANOĞLU, K.; YESILADA, Y. Web accessibility in turkey: awareness, understanding and practices of user experience professionals. *Universal Access in the Information Society*, v. 18, n. 2, p. 387–398, Jun 2019. ISSN 1615-5297. Disponível em: <https://doi.org/10.1007/s10209-017-0603-3>. Citado na página 8.

ISO. *ISO/IEC 25010:2011: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models*. Geneva, CH, 2011. Disponível em: <https://www.iso.org/standard/35733.html>. Citado na página 7.

ISO. *ISO 9241: Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*. Geneva, CH, 2018. Disponível em: <https://www.iso.org/standard/63500.html>. Citado na página 7.

LAZAR, J. Web accessibility policy and law. In: _____. *Web Accessibility: A Foundation for Research*. London: Springer London, 2019. p. 247–261. ISBN 978-1-4471-7440-0. Disponível em: https://doi.org/10.1007/978-1-4471-7440-0_14. Citado na página 7.

LAZAR, J.; DUDLEY-SPONAUGLE, A.; GREENIDGE, K.-D. Improving web accessibility: a study of webmaster perceptions. *Computers in human behavior*, Elsevier, v. 20, n. 2, p. 269–288, 2004. Citado na página 8.

LEITE, M. *Um estudo sobre o conhecimento em acessibilidade digital entre desenvolvedores de aplicações móveis no Brasil*. 2020. Nenhuma citação no texto.

LI, X.; ZHANG, Z.; STEFANIDIS, K. Mobile app evolution analysis based on user reviews. In: *SoMeT*. [S.l.: s.n.], 2018. Citado 4 vezes nas páginas 8, 9, 24 e 33.

MARTIN, W.; SARRO, F.; JIA, Y.; ZHANG, Y.; HARMAN, M. A survey of app store analysis for software engineering. *IEEE Transactions on Software Engineering*, IEEE, v. 43, n. 9, p. 817–847, 2017. Nenhuma citação no texto.

MCILROY, S.; ALI, N.; KHALID, H.; HASSAN, A. E. Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews. *Empirical Softw. Engg.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 21, n. 3, p. 1067–1106, jun. 2016. ISSN 1382-3256. Disponível em: <http://dx.doi.org/10.1007/s10664-015-9375-7>. Citado 4 vezes nas páginas 9, 23, 24 e 30.

MELO, A. M.; BARANAUSKAS, M. C. C.; BONILHA, F. F. G. Avaliação de acessibilidade na web com a participação do usuário: um estudo de caso. p. 181 – 184, 2004. Nenhuma citação no texto.

NAYEBI, M.; ADAMS, B.; RUHE, G. Release practices for mobile apps – what do users and developers think? v. 1, p. 552 – 562, 2016. Citado na página 8.

Oliveira, A. D. A.; Eler, M. M. Strategies and challenges on the accessibility and interoperability of e-government web portals: A case study on brazilian federal universities. In: *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*. [S.l.: s.n.], 2017. v. 1, p. 737–742. ISSN 0730-3157. Citado na página 8.

ORTEGA, A. E. G. *Mining User Reviews from Mobile Applications for Software Evolution*. Tese (Doutorado) — Technical University of Munich, 2015. Citado 2 vezes nas páginas 8 e 9.

Pagano, D.; Maalej, W. User feedback in the appstore: An empirical study. In: *2013 21st IEEE International Requirements Engineering Conference (RE)*. [S.l.: s.n.], 2013. p. 125–134. ISSN 1090-705X. Citado 3 vezes nas páginas 9, 23 e 24.

Palomba, F.; Linares-Vásquez, M.; Bavota, G.; Oliveto, R.; Di Penta, M.; Shyrovanyk, D.; De Lucia, A. User reviews matter! tracking crowdsourced reviews to support evolution of successful apps. In: *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. [S.l.: s.n.], 2015. p. 291–300. Citado 3 vezes nas páginas 8, 24 e 32.

PALOMBA, F.; LINARES-VÁSQUEZ, M.; BAVOTA, G.; OLIVETO, R.; PENTA, M. D.; POSHYVANYK, D.; LUCIA, A. D. Crowdsourcing user reviews to support the evolution of mobile apps. *Journal of Systems and Software*, v. 137, p. 143 – 162, 2018. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0164121217302807>. Citado 4 vezes nas páginas 8, 24, 33 e 54.

Panichella, S.; Di Sorbo, A.; Guzman, E.; Visaggio, C. A.; Canfora, G.; Gall, H. C. How can i improve my app? classifying user reviews for software maintenance and evolution. In: *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. [S.l.: s.n.], 2015. p. 281–290. Citado 4 vezes nas páginas 9, 24, 25 e 54.

PASSERINO, S. P. M. L. M. Inclusão social via acessibilidade digital: proposta de inclusão digital para pessoas com necessidades especiais. abr. 2007. Disponível em: <https://doi.org/10.30962/ec.144>. Nenhuma citação no texto.

Pelloni, L.; Grano, G.; Ciurumelea, A.; Panichella, S.; Palomba, F.; Gall, H. C. Becloma: Augmenting stack traces with user review information. In: *2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. [S.l.: s.n.], 2018. p. 522–526. Citado 3 vezes nas páginas 9, 24 e 29.

PUTNAM, C.; WOZNIAK, K.; ZEFELDT, M. J.; CHENG, J.; CAPUTO, M.; DUFFIELD, C. How do professionals who create computing technologies consider accessibility? In: *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*. New York, NY, USA: ACM, 2012. (ASSETS '12), p. 87–94. ISBN 978-1-4503-1321-6. Disponível em: <http://doi.acm.org/10.1145/2384916.2384932>. Citado na página 8.

SANTAROSA, L. M. C. *Inclusão digital: espaço possível para pessoas com necessidades educacionais especiais*. 2002. 13-30 p. Nenhuma citação no texto.

SERRA, L. C.; CARVALHO, L. P.; FERREIRA, L. P.; VAZ, J. B. S.; FREIRE, A. P. Accessibility evaluation of e-government mobile applications in brazil. *Procedia Computer Science*, v. 67, p. 348 – 357, 2015. ISSN 1877-0509. Proceedings of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1877050915031257>. Citado na página 8.

SHANLEY, C. *Cracking Accessibility on Mobile Devices: The definitive field guide to accessibility and digital inclusion for business managers and project teams*. 1. ed. [S.l.]: RS Books, 2016. Nenhuma citação no texto.

SILVA, C. *Um estudo sobre a automação de testes de acessibilidade em dispositivos móveis: uma análise de recomendações para de ciência visual da BBC*. 2020. Citado 2 vezes nas páginas 12 e 20.

SILVA, C.; ELER, M. M.; FRASER, G. A survey on the tool support for the automatic evaluation of mobile accessibility. In: *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*. New York, NY, USA: ACM, 2018. (DSAI 2018), p. 286–293. ISBN 978-1-4503-6467-6. Disponível em: <http://doi.acm.org/10.1145/3218585.3218673>. Citado na página 7.

TAMJEED, M. *Accessibility in User Reviews for Mobile Apps: An Automated Detection Approach*. 2020. Citado 2 vezes nas páginas 23 e 34.

Vendome, C.; Solano, D.; Linan, S.; Linares-Vasquez, M. Can everyone use my app? an empirical study on accessibility in android apps. In: *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. [S.l.: s.n.], 2019. p. 41–52. Citado na página 8.

W3C. *Accessibility Fundamentals – Introduction to Web Accessibility*. 2018. Disponível em: <https://www.w3.org/standards/webdesign/accessibility>. Citado 2 vezes nas páginas 7 e 12.

W3C. *Web Content Accessibility Guidelines (WCAG) 2.1*. 2018. Disponível em: <https://www.w3.org/TR/WCAG21/>. Citado 3 vezes nas páginas 7, 13 e 37.

W3C. *Making the Web Accessible*. 2020. Disponível em: <https://www.w3.org/WAI/>. Nenhuma citação no texto.

YAN, S.; RAMACHANDRAN, P. G. The current status of accessibility in mobile apps. *ACM Trans. Access. Comput.*, v. 12, n. 1, p. 3:1–3:31, fev. 2019. ISSN 1936-7228. Disponível em: <http://doi.acm.org/10.1145/3300176>. Citado na página 8.